QLOGIC®

The Ultimate in Performance

Installation Guide

Storage Networking (Unified Fabric Pilot)

QLogic 8100 Series Adapters, QLogic 5800 Series Switches, Cisco Nexus 5000 Series Switches, JDSU Xgig Platform

51031-00 A

Information furnished in this manual is believed to be accurate and reliable. However, QLogic Corporation assumes no responsibility for its use, nor for any infringements of patents or other rights of third parties which may result from its use. QLogic Corporation reserves the right to change product specifications at any time without notice. Applications described in this document for any of these products are for illustrative purposes only. QLogic Corporation makes no representation nor warranty that such applications are suitable for the specified use without further testing or modification. QLogic Corporation assumes no responsibility for any errors that may appear in this document.

Document Revision History	
Revision A, August, 2010	
Changes	**Sections Affected**

Table of Contents

Preface

Intended Audience vii
Related Materials vii
Documentation Conventions viii

1 Overview

2 Planning

Selecting a Test Architecture 2-1
Organizational Ownership—Fibre Channel/Storage, Ethernet/Networking . 2-1
Where and How to Deploy 2-1

3 Architecture

Approach 3-1
Process Summary 3-1
Reference Architecture Description 3-1
Architecture Overview 3-1
Equipment Details 3-3

4 Installation

Determine Configuration 4-1
Equipment Installation and Configuration. 4-2
Install the Converged Network Adapter Hardware 4-2
Install the Adapter Drivers 4-3
Install SANsurfer FC HBA Manager 4-4
Cabling 4-5
Fibre Channel Switches. 4-5
FCoE Switches 4-5
Configuring DCB on a Nexus Switch. 4-6
Storage 4-9
Verify Equipment Connectivity 4-10

5 Validation Methods for DCB and FCoE

Validation Step 1—Verify PFC and ETS Parameters 5-1
Validation Process 5-1

Validation Results 5-4
Validation Step 2—FIP Validation 5-6
Validation Process 5-7
Validation Results 5-8
Validation Step 3—FCoE Function and I/O Tests 5-12
Validation Process 5-12
Validation Results 5-13
Validation Step 4—Validate PFC and Lossless Link 5-15
Validation Process 5-15
Validation Results 5-16
Validation Step 5—ETS Verification 5-18
Validation Process 5-18
Validation Results 5-20
Exchange Completion Time (Latency Monitor) 5-25
Validation Step 6—iSCSI Function and I/O Test 5-26
Validation Process 5-26
Validation Results 5-27
Validation Step 7—Virtualization Verification 5-29
Validation Process 5-29
Validation Results 5-30

A Hardware and Software

Cisco Unified Fabric Switch A-1
JDSU (formerly Finisar) Equipment and Software A-2
JDSU Xgig A-2
JDSU Medusa Labs Test Tool Suite A-2
QLogic QLE8142 Converged Network Adapter A-3
QLogic 5800V Series Fibre Channel Switch A-3

B Data Center Bridging Technology

Data Center Bridging (DCB) B-1
DCBX and ETS B-1
Priority Flow Control B-2
Fibre Channel Over Ethernet B-3
iSCSI B-4

C References

Index

List of Figures

Figure		Page
3-1	Reference Architecture Diagram	3-2
4-1	Driver Download Page Model Selection	4-3
4-2	Driver Download Page Driver Link	4-4
4-3	Driver Download Page Driver and Documentation	4-4
4-4	Device Manager Interface—Verify Server and Adapter Login	4-10
4-5	Device Manager Interface—Verify Storage Login	4-10
4-6	NETApp Zone Validation	4-11
4-7	SANsurfer Management Validation Screen 1	4-12
4-8	SANsurfer Management Validation Screen 2	4-12
5-1	Setup for Validating Unified Fabric	5-1
5-2	Analyzer TraceControl Configured to Capture LLDP Frames Only Between the Adapter, Switch, and Target	5-2
5-3	DCBX Exchanges Between QLogic QLE8152 Adapter and the Cisco Nexus 5000 FCoE Switch	5-4
5-4	FCoE Switch Validation with Emulator Setup	5-5
5-5	FIP Test Setup	5-6
5-6	FIP Verification Results—Converged Network Adapter and FCoE Switch	5-8
5-7	Test Results of FIP Keep_Alive and Discovery Advertisement for FCoE Virtual Link Maintenance	5-9
5-8	Time Difference Between Two Adjacent FIP Keep_Alive Frames	5-9
5-9	Time Difference Between Two Adjacent FIP Discovery Advertisement Multicast Frames (to All_E_Nodes)	5-10
5-10	Link De-instantiated by the Switch 20 Seconds after the Last Keep Alive Frame	5-10
5-11	Time Difference Between the Last FIP Keep Alive Sent and the FIP Clear Virtual Link Message from the Switch	5-11
5-12	I/O Test with Fibre Channel SAN Storage Setup	5-12
5-13	MLTT Configures Various I/O Applications to Verify I/O Benchmarking Performance of Different Storage Networks	5-13
5-14	PFC Validation Setup	5-15
5-15	Trace Shows the PFC Event to Pause the Transmitting FCoE Data Frame Further	5-16
5-16	Expert Validating the Reaction to PFC Pause Frames	5-17
5-17	Expert Generates Reports with Detailed Protocol Analysis and Statistics	5-17
5-18	Verifying ETS with Write Operations to the Storage	5-19
5-19	Verifying ETS with Read Operations from the Storage	5-19
5-20	Throughput per Priority	5-20
5-21	FCoE Traffic in a Read Operation from Storage	5-21
5-22	FCoE Traffic Plus TCP Traffic in a Read from Storage	5-21
5-23	Enabling Cross Port Analysis in Expert	5-22
5-24	Switch Issue Pause to Storage	5-22
5-25	LAN Traffic Only	5-23
5-26	Relationship Between FCoE, TCP Traffic, and Pause Frames	5-24
5-27	Read Latency Measurement	5-25

5-28 iSCSI Traffic Performance Validation Setup 5-26
5-29 Throughput Results Showing the Fibre Channel and iSCSI Application Performance 5-27
5-30 Setup for Verifying Virtualization 5-29
5-31 FCoE and iSCSI Application Throughput in a Virtual Environment 5-30
5-32 Expert Shows PFC Pause Request Released from the Target to the Switch Port. . 5-31
B-1 Priority Flow Control B-3
B-2 FCoE Mapping Illustration (Source: FC-BB-5 Rev 2.0) B-3

List of Tables

Table **Page**

3-1 Converged Network Equipment Inventory 3-3
3-2 JDSU Converged Network Validation Equipment 3-3
5-1 I/O Performance Comparison Among FCoE Storage Network Scenarios 5-14
5-2 FCoE Application Performance Statistics Evaluated by Expert 5-18
5-3 Analyzer Port Setup Reference for Monitoring Links 5-20
5-4 Summary of Throughput Characteristic 5-23
5-5 Performance Comparison Between Separate and Combined Applications 5-28
5-6 FCoE and iSCSI Application Throughput when Running Separate Virtual Machines on One Physical Server 5-31

Preface

This document describes how to install and certify a pilot unified fabric configuration. This configuration demonstrates lossless Ethernet and data center bridging (DCB), which includes priority flow control (PFC), enhanced transmission selection (ETS), and data center bridging exchange protocol (DCBX) for a Fibre Channel and 10Gb Ethernet unified fabric.

Intended Audience

This guide is for system engineers and planners who want to provide converged networking products, solutions, and services to their customers. It is also intended for network planners and administrators who are implementing a converged network for their company. This guide describes how to install and validate a pilot converged network in preparation for production deployment.

This guide assumes basic knowledge of Enhanced Ethernet and the associated standards. If the you are not familiar with Fibre Channel over Ethernet (FCoE) and Enhanced Ethernet, review the documents listed in “Related Materials” on page vii.

Related Materials

In addition to this guide, the *FCoE Design Guide,* developed jointly by Cisco and QLogic (Cisco document number C11-569320-01), is a valuable resource and is referenced throughout this guide.

The following links provide more detailed information, and connect to the IEEE documents that define the Enhanced Ethernet functions:

- P802.1Qbb: Priority-based Flow Control
  http://www.ieee802.org/1/files/public/docs2008/bb-pelissier-pfc-proposal-0508.pdf
- P802.1Qaz: Enhanced Transmission Selection (Priority Groups):
  http://www.ieee802.org/1/files/public/docs2008/az-wadekar-ets-proposal-0608-v1.01.pdf
- P802.1Qaz: DCB Capability Exchange Protocol (DCBX):

  http://www.ieee802.org/1/files/public/docs2008/az-wadekar-dcbx-capability-exchange-discoveryprotocol-1108-v1.01.pdf

The Ethernet Alliance has white papers that further describe Enhanced Ethernet: http://www.ethernetalliance.org/library/ethernet_in_the_data_center/white_papers

Documentation Conventions

This guide uses the following documentation conventions:

- **NOTE:** provides additional information.
- ***CAUTION!*** indicates the presence of a hazard that has the potential of causing damage to data or equipment.
- ***WARNING!!*** indicates the presence of a hazard that has the potential of causing personal injury.
- Text in blue font indicates a hyperlink (jump) to a figure, table, or section in this guide, and links to Web sites are shown in underlined blue. For example:
  - Table 9-2 lists problems related to the user interface and remote agent.
  - See "Installation Checklist" on page 3-6.
  - For more information, visit www.qlogic.com.
- Text in **bold** font indicates user interface elements such as a menu items, buttons, check boxes, or column headings. For example:
  - Click the **Start** button, point to **Programs**, point to **Accessories**, and then click **Command Prompt**.
  - Under **Notification Options**, select the **Warning Alarms** check box.
- Text in `Courier` font indicates a file name, directory path, or command line text. For example:
  - To return to the root directory from anywhere in the file structure: Type `cd /root` and press ENTER.
  - Enter the following command: `sh ./install.bin`
- Key names and key strokes are indicated with UPPERCASE:
  - Press CTRL+P.
  - Press the UP ARROW key.

- Text in *italics* indicates terms, emphasis, variables, or document titles. For example:
  - For a complete listing of license agreements, refer to the *QLogic Software End User License Agreement*.
  - What are *shortcut keys*?
  - To enter the date type *mm/dd/yyyy* (where *mm* is the month, *dd* is the day, and *yyyy* is the year).
- Topic titles between quotation marks identify related topics either within this manual or in the online help, which is also referred to as *the help system* throughout this document.

1 Overview

Martin (2010) defines a converged network as follows: "A unified data center fabric is a networking fabric that combines traditional LAN and storage area network (SAN) traffic on the same physical network with the aim of reducing architecture complexity and enhancing data flow and access. To make this work, the traditional Ethernet network must be upgraded to become *lossless* and provide additional data center networking features and functions. In turn, the storage protocol must be altered to run on Ethernet." Lossless means that no Fibre Channel packets are dropped.

Deploying FCoE over a unified fabric reduces data center costs by converging data and storage networking. Standard TCP/IP and Fibre Channel traffic share the same high-speed 10Gbps Ethernet wire, resulting in cost savings through reduced adapter, switch, cabling, power, cooling, and management requirements. FCoE has rapidly gained market acceptance because it delivers excellent performance, reduces data center total cost of ownership (TCO), and protects current data center investments. A unified fabric with FCoE preserves existing investments in Fibre Channel and Ethernet while providing Enhanced Ethernet for unified data networking.

To provide some assurance that FCoE can be deployed in a data center without disrupting operations, Cisco®, JDSU™ (formerly Finisar®), and QLogic have collaborated to produce this guide to simplify the installation and certification of an FCoE pilot. The guide provides system engineers, architects, and end users with a step-by-step method to implement and validate a unified fabric and measure performance of a pilot operation. This guide does not provide methods to measure performance under load or to contrast performance between various protocols, media types, or file systems. This guide is intended to assist in implementing an FCoE and unified fabric pilot using current storage and protocols. It is also intended to assist system engineers in implementing an FCoE and unified fabric pilot for their customers.

2 Planning

Selecting a Test Architecture

When planning the pilot of a unified network, it is important to choose both Fibre Channel and traditional Ethernet-based traffic flows. Combining a test SAN infrastructure and a test LAN infrastructure is often the easiest and most available option for a pilot project. Alternatively, a critical business application test system can closely simulate a production environment. The architecture you choose must demonstrate that a unified network improves efficiency and performance in your environment. The reference architecture described in this guide was assembled from equipment that was available in the QLogic NETtrack Developer Center. You will need to substitute your own equipment, and modify the installation and validation process accordingly.

Organizational Ownership—Fibre Channel/Storage, Ethernet/Networking

A critical factor for successfully implementing a unified data center fabric is the stability of network and storage management practices. Cooperation between the network and storage management teams is important as they configure the converged data center fabric.

Where and How to Deploy

The unified fabric has two components:

- 10Gb Ethernet switches that support DCB and FCoE—These switches support the connection of traditional Ethernet and Fibre Channel infrastructures. These switches are known as a top of rack (TOR) switches, implementing DCB and encapsulating Fibre Channel frames into Ethernet frames for transport over 10Gb Ethernet media.
- 10Gb Converged Network Adapters that support both Ethernet LAN and Fibre Channel SAN over 10Gb Ethernet media—These adapters replace the NIC and Fibre Channel host bus adapter, and connect to a DCB-enabled 10Gb Ethernet switch.

Currently, a Converged Network Adapter must always be connected to a switch that has DCB. There are two types of switches that have DCB: a DCB switch and an FCoE switch. The DCB switch has enhanced Ethernet support, but does not have Fibre Channel forwarder (FCF) capabilities, and does not support the conversion of Fibre Channel frames to FCoE frames. A DCB switch supports converging-Ethernet-based protocols, but does not support Fibre Channel protocols. The DCB switch requires an external device to manage Fibre Channel and FCoE functions. An FCoE switch supports both DCB and Fibre Channel.

There are three ways to connect Fibre Channel storage to a unified fabric:

- **Converged Network Adapter > FCoE switch > Fibre Channel switch > Fibre Channel storage**

  The adapter connects to the FCoE switch with Ethernet infrastructure, and the FCoE switch connects to storage through a Fibre Channel switch (Martin, 2010). This is the most common implementation in today's data centers because the Fibre Channel switch and SAN storage are typically already in place.

- **Converged Network Adapter > DCB switch > FCF > Fibre Channel switch > Fibre Channel storage**

  The DCB switch requires an external device to provide FCF function to the attached Fibre Channel storage. This approach is not as common because most data centers do not have an FCF device, and will acquire an FCoE switch to connect to their Fibre Channel Infrastructure.

- **Converged Network Adapter > FCoE switch > FCoE storage**

  This implementation is not common because, most data centers use Fibre Channel SAN storage. As more storage vendors deliver FCoE storage, more pilot projects will support direct Ethernet connection from the FCoE switch to FCoE-capable storage controllers (Martin, 2010).

In all cases, Ethernet LAN and iSCSI storage connect directly to Ethernet ports on the DCB or FCoE switch.

The reference architecture, described in Section 3, uses direct-attached native FCoE storage and Fibre Channel switch-connected SAN storage. Section 4 describes the implementation of the reference architecture and Section 5 describes the validation.

3 Architecture

Approach

The test configuration described in this section was installed and validated at the QLogic NETtrack Developer Center (NDC) located in Shakopee, MN. The NDC provides QLogic customers and alliance partners with the test tracks and engineering staff to test interoperability and optimize performance with the latest server, networking, and storage technology.

Process Summary

To establish a repeatable process, the team created a converged network in the NDC and installed a validation environment based on JDSU hardware and software. Screen shots and trace data were captured to show the results from the seven-step validation process, and to demonstrate the various DCB functions.

Reference Architecture Description

Architecture Overview

Figure 3-1 illustrates the unified Ethernet infrastructure enhanced with DCB. FCoE and iSCSI storage traffic was tested and validated with LAN traffic, which shared the unified 10GbE bandwidth driven by Converged Network Adapters. JDSU testing tools were installed in the fabric, either in-line or at the edge, with hardware and software to test and certify system-level behavior. These tools also provided expert-system support that simplified troubleshooting and reduced installation time.

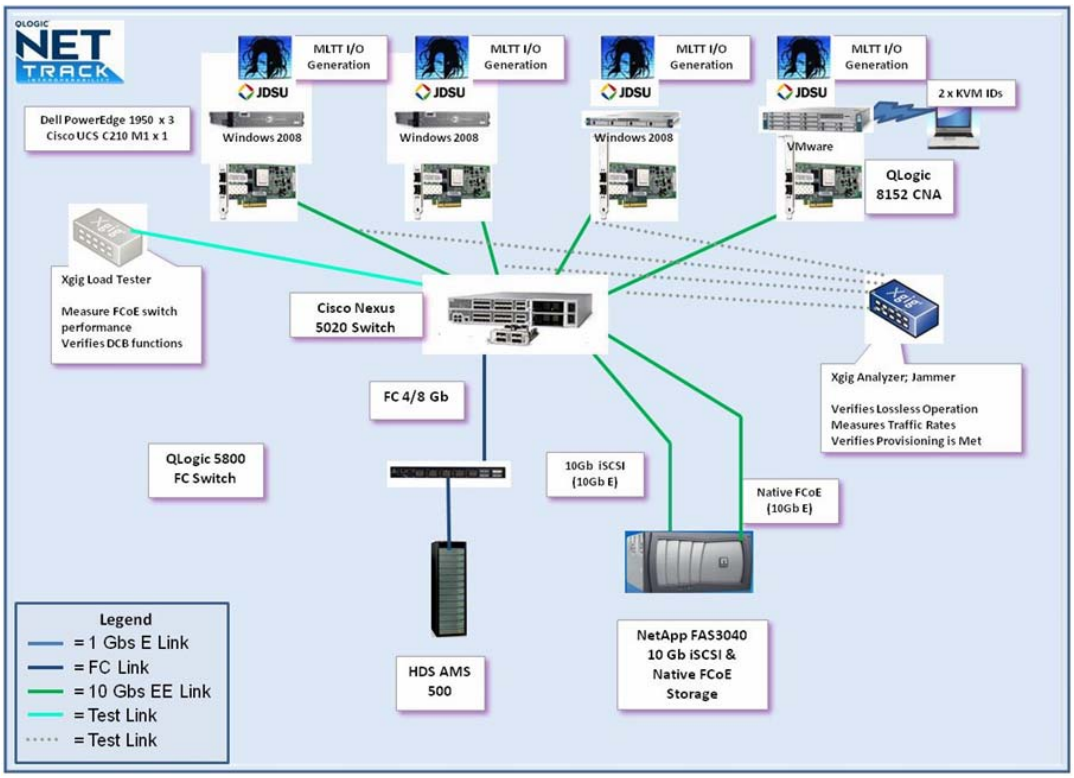


Figure 3-1. Reference Architecture Diagram

Equipment Details

Table 3-1 lists the referenced architecture equipment. Table 3-2 lists the JDSU Xgig equipment and testing software.

Table 3-1. Converged Network Equipment Inventory

Quantity	Product	Model Number
3	Dell® PowerEdge® servers [a]	1950
4	QLogic QLE8142 adapters (server connectivity)	QLE8142
1	Cisco UCS general-purpose rack-mount server Windows® 2008	UCS C210 M1
1	Cisco UCS high-density rack-mount server Windows 2008	UCS C200 M1
1	NetApp® FAS3040 storage system (10Gb iSCSI storage and 10Gb native FCoE storage)	FAS3040
1	HDS® AMS500 Storage Array (4Gb Fibre Channel storage)	AMS500
1	QLogic 5800V 8Gb Fibre Channel switch (Fibre Channel connection)	SB5802V
1	Cisco Nexus™ 5020 FCoE switch (unified fabric interconnect)	Nexus 5020

[a] Two Windows 2008, one VMware® ESX™ with one Windows 2008 and Red Hat® Enterprise Linux guest with Microsoft® iSCSI initiator. All servers use the QLogic QLE8142 adapter.

Table 3-2. JDSU Converged Network Validation Equipment

Product	Model/Part Number
Xgig four-slot chassis	Xgig-C042
10GbE/10G Fibre Channel multi-functional blade	Xgig-B2100C
Four-port, 1, 2, 4, 8Gbps Fibre Channel Analyzer blade	Xgig-B480FA
Two-port, Xgig 10GbE Fibre Channel Analyzer function key	Xgig-2FG10G1-SW
Two-port, 10GE Analyzer and Jammer function key	Xgig-S20JFE
Two-port, 10GEE FCoE Analyzer and Load Tester function key	Xgig-S20LEA
Four-port, 1, 2, 4, 8Gbps Fibre Channel Analyzer function key	Xgig-S48AFA

4 Installation

This section describes how to set up an FCoE environment. It assumes a general understanding of SAN administration concepts.

Determine Configuration

QLogic FCoE adapters are supported on multiple hardware platforms and operating systems. Generally, the following specifications apply, but you should always check the QLogic Web site for current information. This configuration uses a subset of the following equipment:

- Server bus interface—PCIe® Gen1 x8 or PCIe Gen2 x4
- Hardware platforms—IA32 (x86), Intel64, AMD64 (x64), IA64, SPARC®, PowerPC®
- Operating systems—Windows Server® 2003, 2008, 2008 R2 (targeted); Red Hat EL AP 5.*x*; Novell® SLES 10.*x*, 11; VMware ESX/ESXi 3.5 and 4.0; Solaris® 10, OpenSolaris™. This list can be found under the specifications tab for QLogic adapters on the QLogic Web site.
- Storage—The following storage systems are in most data centers:
  - Fibre Channel
  - iSCSI
  - FCoE (NetApp)
- Switches—The following switches are typical in this configuration:
  - Fibre Channel
  - FCoE
  - Ethernet
- Cabling
  - Fiber optic cable (OM2/OM3) between servers, switches, and storage
  - Cat5e and Cat6 Ethernet for device management and 1GbE iSCSI storage

Equipment Installation and Configuration

This section focuses on the converged network installation and configuration. You do not have to change your current storage and network management practices.

Install the Converged Network Adapter Hardware

Begin by identifying a pilot server that meets Converged Network Adapter hardware requirements (PCI slot type, length, available slot) and install the adapters.

To install the adapter hardware:

1. Use a ground strap to avoid damaging the card or server.
2. Power off the computer and disconnect the power cable.
3. Remove the computer cover, and find an empty PCIe x8 bus slot (Gen1) or PCIe x4 bus slot (Gen2).
4. Pull out the slot cover (if any) by removing the screw or releasing the lever.
5. Install the low-profile bracket, if required.
6. Grasp the adapter by the top edge, and then insert it firmly into the appropriate slot.
7. Refasten the adapter's retaining bracket using the existing screw or lever.
8. Close the computer cover.
9. Plug the appropriate Ethernet cable (either copper or optical) into the adapter.
   - Optical models ship with optical transceivers installed. QLogic 8100 Series adapters used for this project operate only with optical transceivers sold by QLogic.
   - The list of approved copper cables is available at the following link: http://www.qlogic.com/Products/ConvergedNetworking/ConvergedNetworkAdapters/Pages/CopperCables.aspx.
10. Plug in the power cable, and turn on the computer.

Install the Adapter Drivers

To install the FCoE and Ethernet drivers:

1. Go to the QLogic Driver Downloads/Documentation page (Figure 4-1) at http://driverdownloads.qlogic.com/QLogicDriverDownloads_UI/default.aspx.

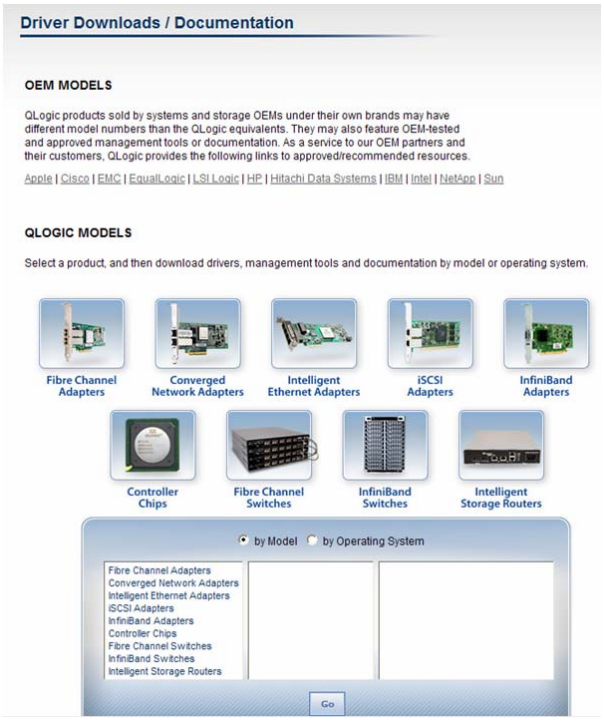

Driver Downloads / Documentation
OEM MODELS
QLogic products sold by systems and storage OEMs under their own brands may have different model numbers than the QLogic equivalents. They may also feature OEM-tested and approved management tools or documentation. As a service to our OEM partners and their customers, QLogic provides the following links to approved/recommended resources.
Apple | Cisco | EMC | EqualLogic | LSI Logic | HP | Hitachi Data Systems | IBM | Intel | NetApp | Sun
QLOGIC MODELS
Select a product, and then download drivers, management tools and documentation by model or operating system.
Fibre Channel Adapters
Converged Network Adapters
Intelligent Ethernet Adapters
iSCSI Adapters
InfiniBand Adapters
Controller Chips
Fibre Channel Switches
InfiniBand Switches
Intelligent Storage Routers
by Model
by Operating System
Fibre Channel Adapters
Converged Network Adapters
Intelligent Ethernet Adapters
iSCSI Adapters
InfiniBand Adapters
Controller Chips
Fibre Channel Switches
InfiniBand Switches
Intelligent Storage Routers
Go

Figure 4-1. Driver Download Page Model Selection

2. In the table at the bottom of the page, select **Converged Network Adapters**, the adapter model, your operating system, and then click **Go**.
3. On the download page under Drivers, select the driver and download it to your system.
4. Follow the included instructions for installing the downloaded driver.

Install SANsurfer FC HBA Manager

To install SANsurfer® FC HBA Manager:

1. Go to the QLogic Driver Downloads/Documentation page at http://driverdownloads.qlogic.com/QLogicDriverDownloads_UI/default.aspx.
2. In the table at the bottom of the page (Figure 4-2), select **Converged Network Adapters**, the adapter model, your operating system, and then click **Go**.

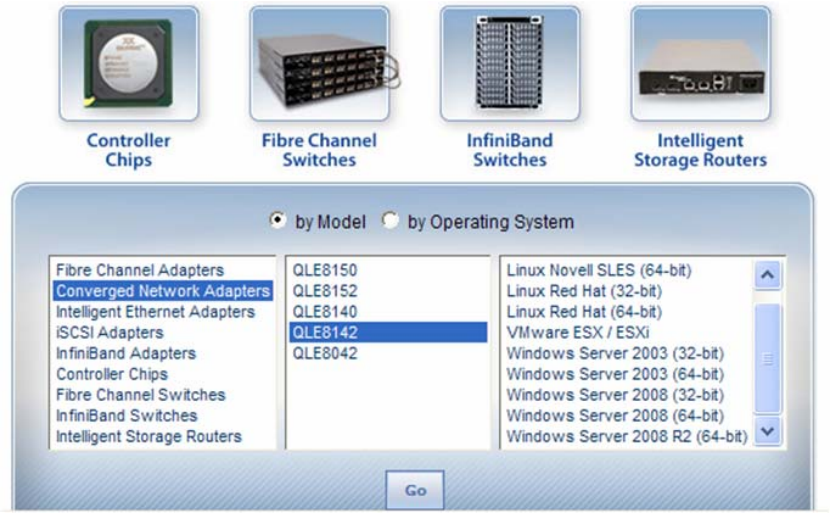


Figure 4-2. Driver Download Page Driver Link

3. On the download page under Management Tools (Figure 4-3), select **SANsurfer FC HBA Manager** and download it to your system.

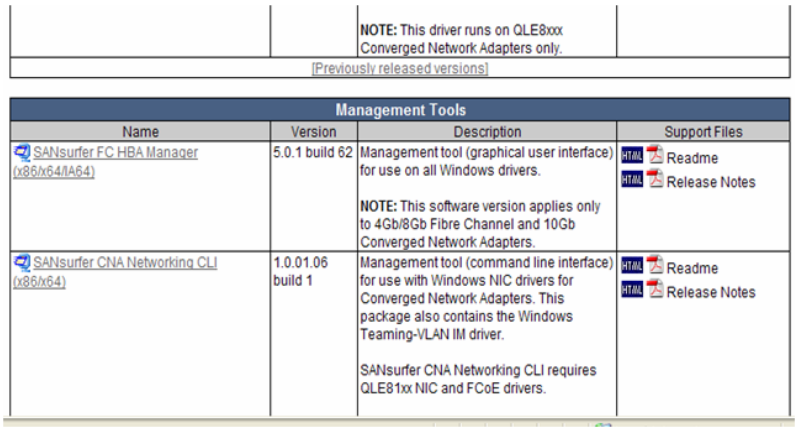

NOTE: This driver runs on QLE8xxx Converged Network Adapters only.
[Previously released versions]
Management Tools
Name
Version
Description
Support Files
SANsurfer FC HBA Manager (x86/x64/IA64)
5.0.1 build 62
Management tool (graphical user interface) for use on all Windows drivers.
NOTE: This software version applies only to 4Gb/8Gb Fibre Channel and 10Gb Converged Network Adapters.
Readme
Release Notes
SANsurfer CNA Networking CLI (x86/x64)
1.0.01.06 build 1
Management tool (command line interface) for use with Windows NIC drivers for Converged Network Adapters. This package also contains the Windows Teaming-VLAN IM driver.
SANsurfer CNA Networking CLI requires QLE81xx NIC and FCoE drivers.
Readme
Release Notes

Figure 4-3. Driver Download Page Driver and Documentation

4. Follow the included instructions for installing the downloaded software.

Cabling

To connect the Fibre Channel and Ethernet cables:

1. Connect the Fibre Channel cables from the servers to the Cisco FCoE Nexus switch.
2. Connect the Fibre Channel cables from the storage to the Cisco FCoE Nexus switch.
3. Connect any necessary Ethernet cables for device management and iSCSI storage.

Fibre Channel Switches

If you are connecting Fibre Channel devices, such as storage, through a Fibre Channel switch, then you must connect the Fibre Channel switch to a Fibre Channel port on the FCoE switch.

In addition, set up a zoning configuration so that the servers can discover the disk LUNs you are mapping. Refer to the Fibre Channel switch documentation for zoning instructions.

FCoE Switches

QLogic and Cisco have jointly developed the *QLogic and Cisco FCoE Design Guide* for implementing a unified data center using Cisco Nexus 5000 Series switches and QLogic second-generation Converged Network Adapters. Refer to the design guide for detailed instructions on how to implement an FCoE network and configure the Cisco Nexus switch and QLogic adapters (Cisco and QLogic, 2010). The design guide also describes how to configure N_Port ID Virtualization (NPIV) to resolve fabric expansion concerns related to domain IDs. The *QLogic and Cisco FCoE Design Guide* does not describe the configuration of the PFC and ETS DCB parameters, which will be required for the tests described in this document.

For information about configuring DCB on a Cisco Nexus 5000 Series switch, see “Configuring DCB on a Nexus Switch” on page 4-6; some variables may need to be adjusted for your configuration.

Configuring DCB on a Nexus Switch

NOTE:

In this procedure, you may need to adjust some of the parameters to suit your environment, such as VLAN IDs, Ethernet interfaces, and virtual Fibre Channel interfaces. In this example, the Cisco FCF uses NIC traffic on priority 2 and VLAN 2, and FCoE traffic on priority 3 and VLAN 1002.

To enable PFC, ETS, and DCB functions on a Cisco Nexus 5000 series switch:

1. Open a terminal configuration setting.

```
Switch# config t
switch(config)#
```

2. Configure qos class-maps and set the traffic priorities: NIC uses priority 0 and FcoE uses priority 3.

```
class-map type qos class-nic
  match cos 0
class-map type qos class-fcoe
  match cos 3
```

3. Configure queuing class-maps.

```
class-map type queuing class-nic
  match qos-group 2
```

4. Configure network-qos class-maps.

```
 class-map type network-qos class-nic
  match qos-group 2
```

5. Configure qos policy-maps.

```
policy-map type qos policy1
    class type qos class-nic
       set qos-group 2
```

6. Configure queuing policy-maps and assign network bandwidth. Divide the network bandwidth evenly between FcoE and NIC traffic.

```
policy-map type queuing policy1
    class type queuing class-nic
       bandwidth percent 50
    class type queuing class-fcoe
       bandwidth percent 50
    class type queuing class-default
       bandwidth percent 0
```

7. Configure network-qos policy maps and set up the PFC for no-drop traffic class.

```
policy-map type network-qos policy1
    class type network-qos class-nic
       pause no-drop
```

8. Apply the new policy (PFC on NIC and FcoE traffic) to the entire system.

```
system qos
       service-policy type qos input policy1
       service-policy type queuing output policy1
       service-policy type queuing input policy1
       service-policy type network-qos policy1
```

9. Create a unique VLAN for FCoE traffic and NIC traffic.

```
Vlan 2
       Exit
Vlan 1002
       Fcoe vsan 1
       Exit
```

10. Configure Ethernet port 1/3 to enable VLAN-1002-tagged FCoE traffic and VLAN-2-tagged NIC traffic.

```
interface Ethernet1/3
  switchport mode trunk
  switchport trunk allowed vlan 2,1002
  spanning-tree port type edge trunk
interface vfc3
  bind interface Ethernet1/3
  no shutdown
```

11. Display the configuration to confirm that it is correct.

```
switch(config)# sh policy-map system
  Type network-qos policy-maps
  ============================
  policy-map type network-qos policy1
    class type network-qos class-nic       match qos-group 2
      pause no-drop
    class type network-qos class-fcoe      match qos-group 1
      pause no-drop
      mtu 2240
    class type network-qos class-default   match qos-group 0
      mtu 1538

  Service-policy (qos) input:   policy1
    policy statistics status:   disabled
    Class-map (qos):   class-nic (match-any)
      Match: cos 0
      set qos-group 2
    Class-map (qos):   class-fcoe (match-any)
      Match: cos 3
      set qos-group 1
    Class-map (qos):   class-default (match-any)
      Match: any
      set qos-group 0

  Service-policy (queuing) input:   policy1
    policy statistics status:   disabled
    Class-map (queuing):   class-nic (match-any)
      Match: qos-group 2
```

```
    bandwidth percent 50
  Class-map (queuing):   class-fcoe (match-any)
    Match: qos-group 1
    bandwidth percent 50
  Class-map (queuing):   class-default (match-any)
    Match: qos-group 0
    bandwidth percent 0

Service-policy (queuing) output:   policy1
  policy statistics status:   disabled
  Class-map (queuing):   class-nic (match-any)
    Match: qos-group 2
    bandwidth percent 50
  Class-map (queuing):   class-fcoe (match-any)
    Match: qos-group 1
    bandwidth percent 50
  Class-map (queuing):   class-default (match-any)
    Match: qos-group 0
    bandwidth percent 0
```

Storage

Depending on your storage, you may connect directly to the Nexus 5000 switch through FCoE, as in the case of NetApp, or through other methods (Fibre Channel, iSCSI, NFS, CIFS). Consult your disk storage documentation for instructions on how to enable FCoE (NetApp) and assign disk storage LUNs.

Verify Equipment Connectivity

To verify that all equipment is logged in and operating properly:

1. Verify LAN management capability on all devices through the associated device management application.
2. Verify that servers and Converged Network Adapters are logged into an FCoE switch under both Ethernet (eth1/16) and Fibre Channel (vfc116) protocols. Figure 4-4 shows the Device Manager interface for the Cisco Nexus 5000 FCoE switch.

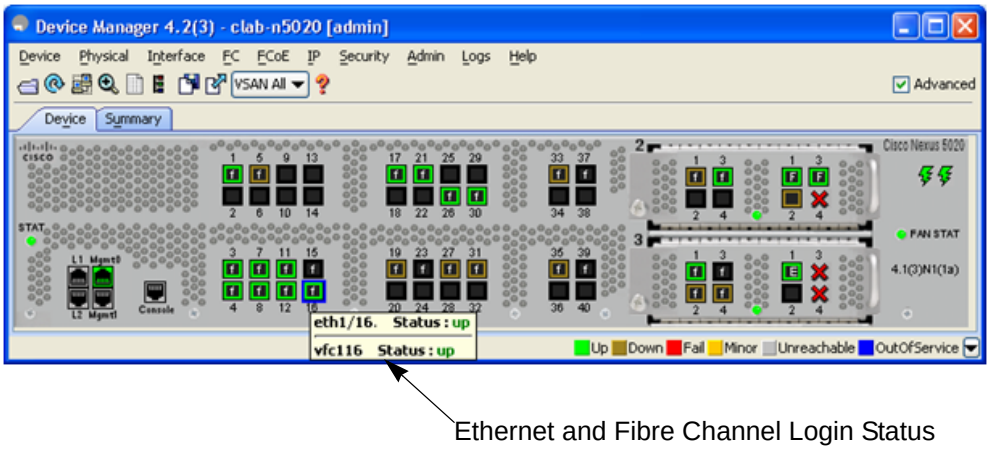


Figure 4-4. Device Manager Interface—Verify Server and Adapter Login

3. Verify that storage devices have logged into a switch (Figure 4-5). Fibre Channel and FCoE storage devices log into Fibre Channel or FCoE switches; iSCSI storage devices log into Ethernet or FCoE switches.

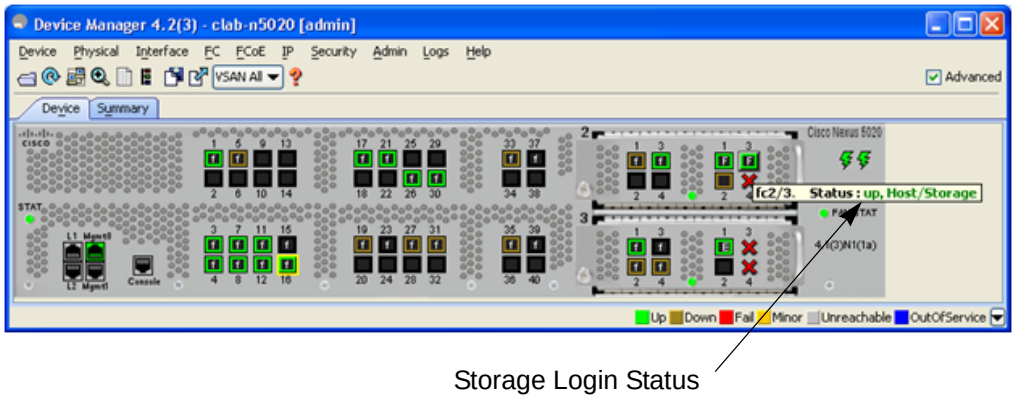

Device Manager 4.2(3) - clab-n5020 [admin]
Device Physical Interface FC FCoE IP Security Admin Logs Help
VSAN All
Advanced
Device Summary
Cisco Nexus 5020
fc2/3. Status : up, Host/Storage
Up Down Fail Minor Unreachable OutOfService
Storage Login Status

Figure 4-5. Device Manager Interface—Verify Storage Login

4. When the LUNs have been created and all zoning is complete, use the management interface to add the Converged Network Adapter Fibre Channel WWNs and iSCSI initiators to your storage so that the servers can discover the LUNs. Figure 4-6 shows an example of the discovered FCoE Converged Network Adapters and the created initiator groups on a NetApp Filer after zoning is complete.

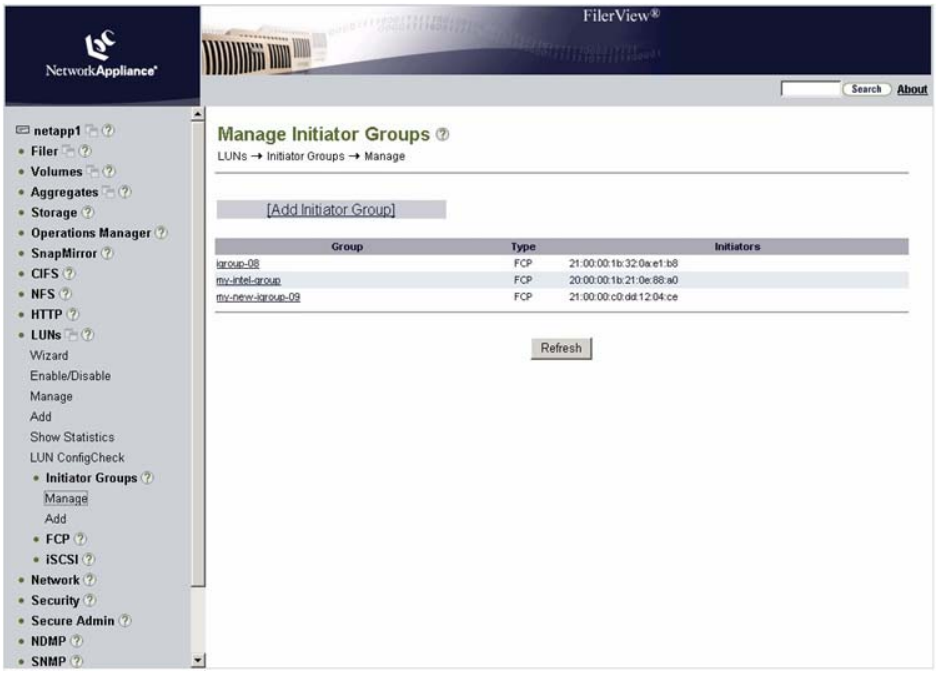


Figure 4-6. NETApp Zone Validation

5. Reboot the servers to discover the assigned LUNs.
6. Verify that server operating system and SANsurfer management application can discover the assigned LUNs (Figures 4-7 and 4-8).

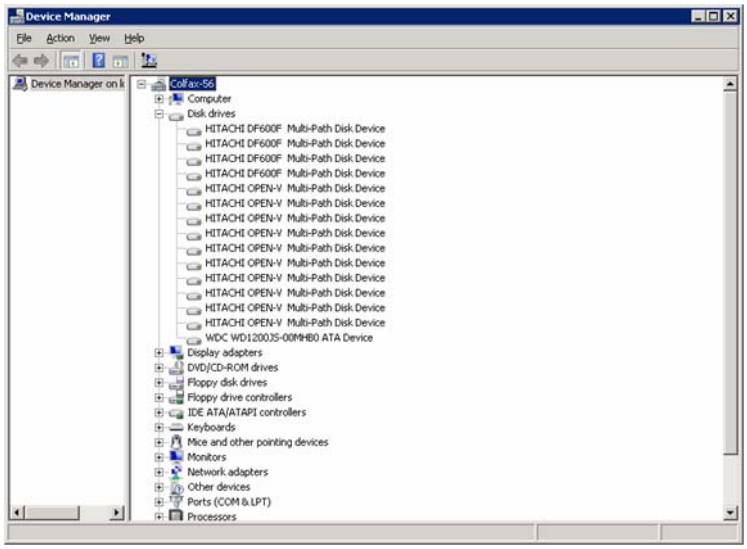


Figure 4-7. SANsurfer Management Validation Screen 1

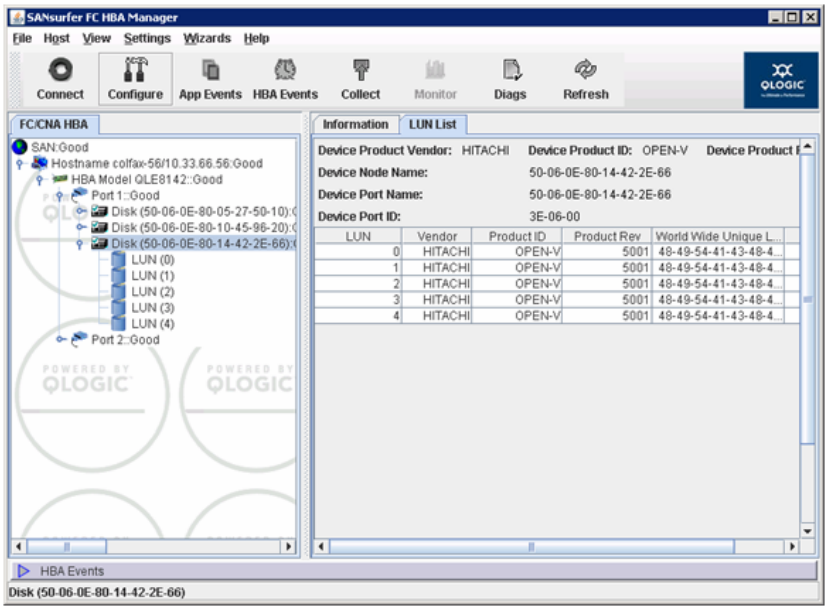


Figure 4-8. SANsurfer Management Validation Screen 2

7. Use the operating system tools to create disk partitions and volumes on your servers using the assigned LUNs.
8. Proceed to validation in Section 5, Validation Methods for DCB and FCoE.

5 Validation Methods for DCB and FCoE

Validation Step 1—Verify PFC and ETS Parameters

Verify the interoperability of the unified fabric devices by capturing the DCBX exchanges between the QLogic initiator and each target.

The key parameters to verify are:

- DCBX version used by each network component
- Priorities assigned to each protocol
- Priorities with the per PFC enabled
- Bandwidth assigned to each priority

Validation Process

1. Place the Analyzer into the two paths shown as in Figure 5-1.

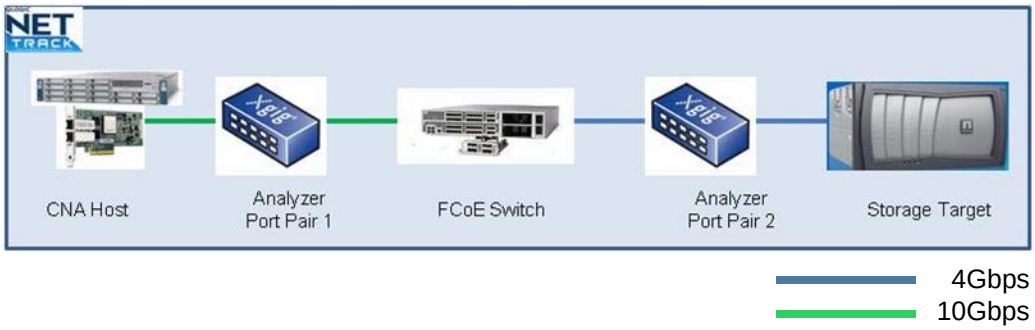


Figure 5-1. Setup for Validating Unified Fabric

2. Configure the two Analyzer port pairs to be in the same time domain to monitor DCBX exchanges between the two links.
3. Configure the Analyzer TraceControl to capture LLDP frames (Figure 5-2):

a. Select the trigger mode.

b. Set up the capture trigger condition to capture any LLDP frames.

c. Set the trigger fill to capture 90 percent after the trigger.

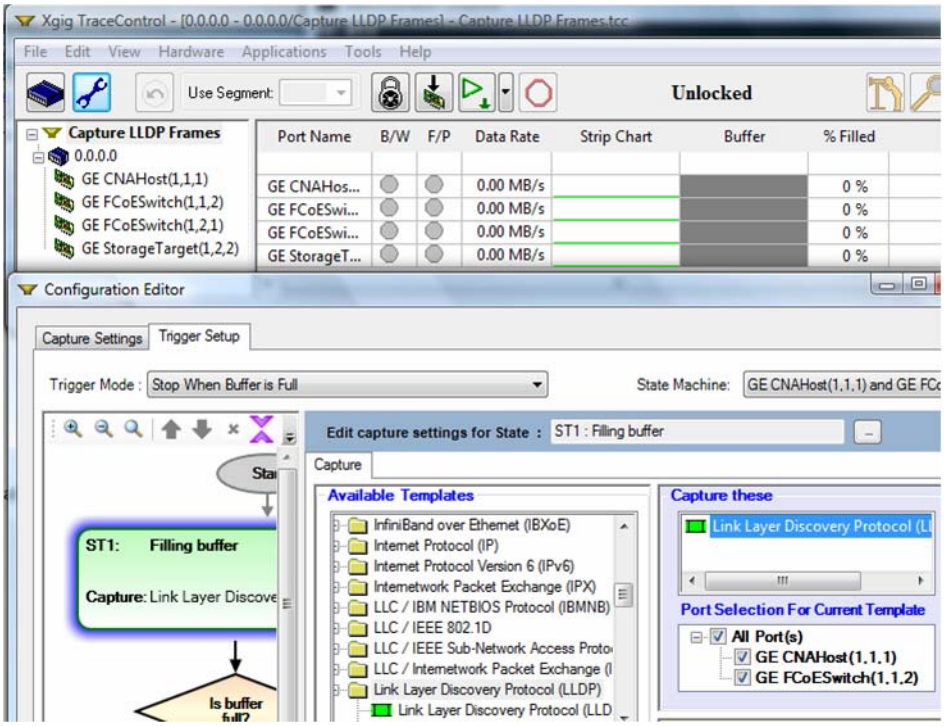

Xgig TraceControl - [0.0.0.0 - 0.0.0.0/Capture LLDP Frames] - Capture LLDP Frames.tcc
File Edit View Hardware Applications Tools Help
Use Segment:
Unlocked
Capture LLDP Frames
0.0.0.0
GE CNAHost(1,1,1)
GE FCoESwitch(1,1,2)
GE FCoESwitch(1,2,1)
GE StorageTarget(1,2,2)
Port Name B/W F/P Data Rate Strip Chart Buffer % Filled
GE CNAHos... 0.00 MB/s 0 %
GE FCoESwi... 0.00 MB/s 0 %
GE FCoESwi... 0.00 MB/s 0 %
GE StorageT... 0.00 MB/s 0 %
Configuration Editor
Capture Settings
Trigger Setup
Trigger Mode : Stop When Buffer is Full
State Machine: GE CNAHost(1,1,1) and GE FC
Edit capture settings for State : ST1 : Filling buffer
Capture
ST1: Filling buffer
Capture: Link Layer Discove
Is buffer
Available Templates
InfiniBand over Ethernet (IBXoE)
Internet Protocol (IP)
Internet Protocol Version 6 (IPv6)
Internetwork Packet Exchange (IPX)
LLC / IBM NETBIOS Protocol (IBMNB)
LLC / IEEE 802.1D
LLC / IEEE Sub-Network Access Proto
LLC / Internetwork Packet Exchange (I
Link Layer Discovery Protocol (LLDP)
Link Layer Discovery Protocol (LLD
Capture these
Link Layer Discovery Protocol (LL
Port Selection For Current Template
All Port(s)
GE CNAHost(1,1,1)
GE FCoESwitch(1,1,2)

Figure 5-2. Analyzer TraceControl Configured to Capture LLDP Frames Only Between the Adapter, Switch, and Target

4. Configure the FCoE switch with the following parameters:
   - FCoE priority 3 PFC enabled, ETS bandwidth = 50 percent
   - All other classes of service lossy, ETS bandwidth = 50 percent
5. Capture the trace and verify the following information:
   - DCBX version used by each link end is the same.
   - If the DCBX version 1.01 is used, then verify the following:
     - DCBX Oper_Version field value is the same in DCBX messages for all devices.
     - ER bit is OFF in each DCBX message.
     - EN bit is ON in each DCBX message.
     - Converged Network Adapter and Target Willing bit (W) is ON, while the Switch W bit is OFF.

- pfc_enable value is the same for the adapter, switch, and targets.
- Application type-length-value (TLV) contains one entry for FCoE protocol and one entry for FCoE initialization protocol (FIP). The user_priority_map should be the same for the adapter, switch, and targets.
- Priority group TLV for the switch contains the ETS bandwidth allocation configured at the switch console.

■ If IEEE DCBX version 1.4 or later is used, then verify the following:

- Converged Network Adapter and Target Willing bit (W) is ON, while the Switch W bit is OFF.
- pfc_enable value is the same for the adapter, switch, and targets.
- Application TLV contains one entry for FCoE protocol and one entry for FIP. The user_priority_map should be the same for the adapter, switch, and targets.
- ETS Recommendation TLV for the switch contains the bandwidth allocation per priority as configured on the switch console.
- ETS Configuration TLV for the adapter and target contain the same bandwidth allocation per priority as the switch.

Validation Results

Figure 5-3 shows one trace captured between the adapter and FCoE switch exchanging DCB parameters using the DCBX protocol. Both peers use version 1.01.

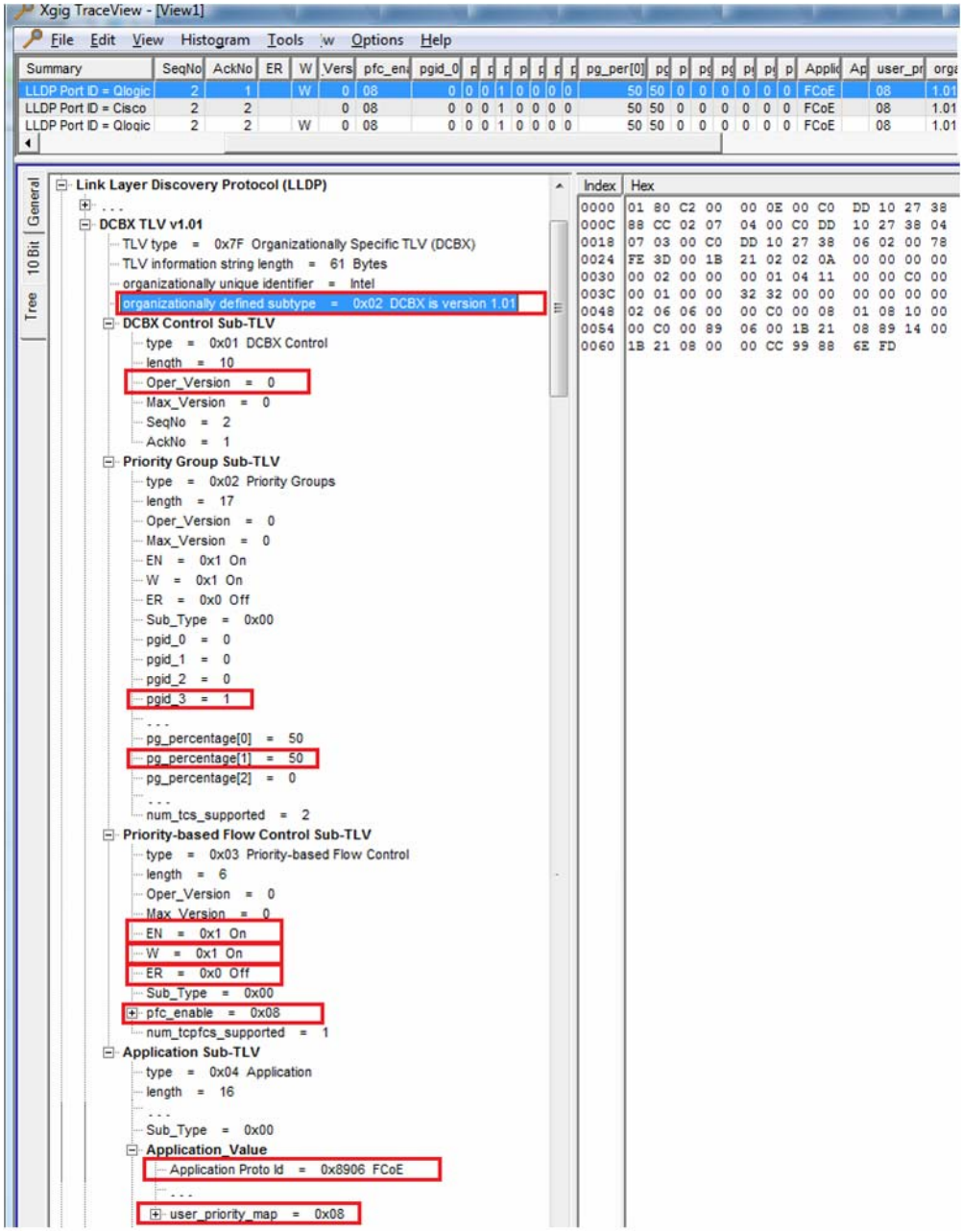

Xgig TraceView - [View1]
File Edit View Histogram Tools Options Help
Summary SeqNo AckNo ER W Vers pfc_ena pgid_0 pg_per[0] Applic user_pr
LLDP Port ID = Qlogic 2 1 W 0 08 0 0 0 1 0 0 0 0 50 50 0 0 0 0 0 0 FCoE 08 1.01
LLDP Port ID = Cisco 2 2 0 08 0 0 0 1 0 0 0 0 50 50 0 0 0 0 0 0 FCoE 08 1.01
LLDP Port ID = Qlogic 2 2 W 0 08 0 0 0 1 0 0 0 0 50 50 0 0 0 0 0 0 FCoE 08 1.01
General
10 Bit
Tree
Link Layer Discovery Protocol (LLDP)
DCBX TLV v1.01
TLV type = 0x7F Organizationally Specific TLV (DCBX)
TLV information string length = 61 Bytes
organizationally unique identifier = Intel
organizationally defined subtype = 0x02 DCBX is version 1.01
DCBX Control Sub-TLV
type = 0x01 DCBX Control
length = 10
Oper_Version = 0
Max_Version = 0
SeqNo = 2
AckNo = 1
Priority Group Sub-TLV
type = 0x02 Priority Groups
length = 17
Oper_Version = 0
Max_Version = 0
EN = 0x1 On
W = 0x1 On
ER = 0x0 Off
Sub_Type = 0x00
pgid_0 = 0
pgid_1 = 0
pgid_2 = 0
pgid_3 = 1
pg_percentage[0] = 50
pg_percentage[1] = 50
pg_percentage[2] = 0
num_tcs_supported = 2
Priority-based Flow Control Sub-TLV
type = 0x03 Priority-based Flow Control
length = 6
Oper_Version = 0
Max_Version = 0
EN = 0x1 On
W = 0x1 On
ER = 0x0 Off
Sub_Type = 0x00
pfc_enable = 0x08
num_tcpfcs_supported = 1
Application Sub-TLV
type = 0x04 Application
length = 16
Sub_Type = 0x00
Application_Value
Application Proto Id = 0x8906 FCoE
user_priority_map = 0x08
Index Hex
0000 01 80 C2 00 00 0E 00 C0 DD 10 27 38
000C 88 CC 02 07 04 00 C0 DD 10 27 38 04
0018 07 03 00 C0 DD 10 27 38 06 02 00 78
0024 FE 3D 00 1B 21 02 02 0A 00 00 00 00
0030 00 02 00 00 00 01 04 11 00 00 C0 00
003C 00 01 00 00 32 32 00 00 00 00 00 00
0048 02 06 06 00 00 C0 00 08 01 08 10 00
0054 00 C0 00 89 06 00 1B 21 08 89 14 00
0060 1B 21 08 00 00 CC 99 88 6E FD

Figure 5-3. DCBX Exchanges Between QLogic QLE8152 Adapter and the Cisco Nexus 5000 FCoE Switch

From the trace, you can verify the following:

- DCBX Oper_Version is 0 for all devices.
- ER bit is OFF in all TLVs for all devices.
- EN bit is ON in all TLVs for all devices messages.
- W bit is ON for the adapter and target, and OFF for the switch.
- pfc_enable value is 0x08 for all devices, which means PFC is enabled for priority 3.
- Application TLV for FCoE contains a user_priority_map of 0x08 for all devices with FCoE traffic running at priority 3.
- Priority Group TLV is the same for the switch and all devices. The pgid_3 (priority 3) is assigned the pg_percentage[1], which has 50 percent of the bandwidth allocation. All other priorities are assigned to pg_percentage[0], and they share the remaining 50 percent of the bandwidth.

If the switch cannot propagate its DCB parameters to the peer devices, then use the JDSU Load Tester to negotiate different versions of the DCBX protocol (Figure 5-4). The Load Tester displays the DCBX settings advertised by the switch at its console. While the Load Tester is negotiating DCBX parameters with the switch, capture the traffic between the Load Tester and the switch using the JDSU Analyzer, and then view the exchanges using TraceView. In TraceView, you can compare the DCBX exchanges with those captured when the switch is connected to the adapter or target.

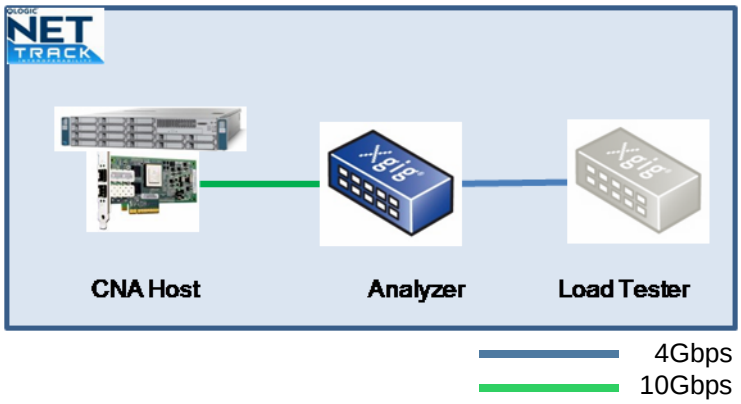


Figure 5-4. FCoE Switch Validation with Emulator Setup

If end devices do not accept the DCB parameters from the switch, configure the same DCB parameters manually on the switches, adapters, and target devices using their configuration consoles. After configuring the parameters, continue to “Validation Step 2—FIP Validation” on page 5-6.

Validation Step 2—FIP Validation

The next step is to verify FCoE initialization protocol (FIP). FIP performs the FCoE link initialization and maintenance to enable FCoE traffic in the unified fabric.

FIP ensures that:

- FCoE switches advertise themselves in broadcast messages.
- End devices query for the VLAN ID to be used for FCoE.
- End devices query for other FCoE characteristics.
- End devices log in to or log out from the switch.
- End devices advertise to the switch that they are present so that the FCoE connection does not timeout.

FIP can be communicated either between the adapter and the FCoE switch (FCF) or between the FCoE target and the FCoE switch (FCF). The FIP test setup is shown in Figure 5-5.

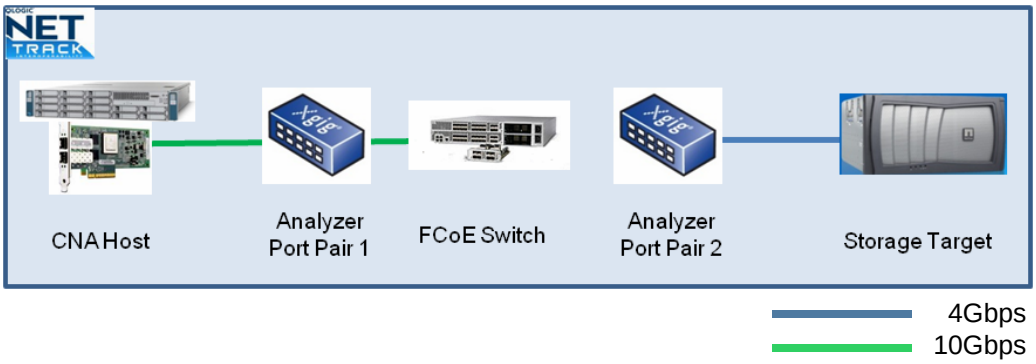


Figure 5-5. FIP Test Setup

Within TraceView, normal FIP operation begins with FIP discovery advertisement broadcast messages from each FCoE switch, which are repeated every eight seconds. These messages contain the MAC address of the switch and the version of the FIP protocol. FIP operation proceeds in the following sequence:

1. The adapter/storage requests the FCoE VLAN ID using the FIP VLAN request message. The 3-bit VLAN priority for FCoE is broadcast through DCBX exchanges every 30 seconds.
2. The FCoE switch responds with the FCoE VLAN ID in the FIP VLAN notification message.
3. The adapters/storage request the FCoE characteristics in an FIP Discovery Solicitation message. This message also transmits the maximum FCoE frame size supported by the end device.

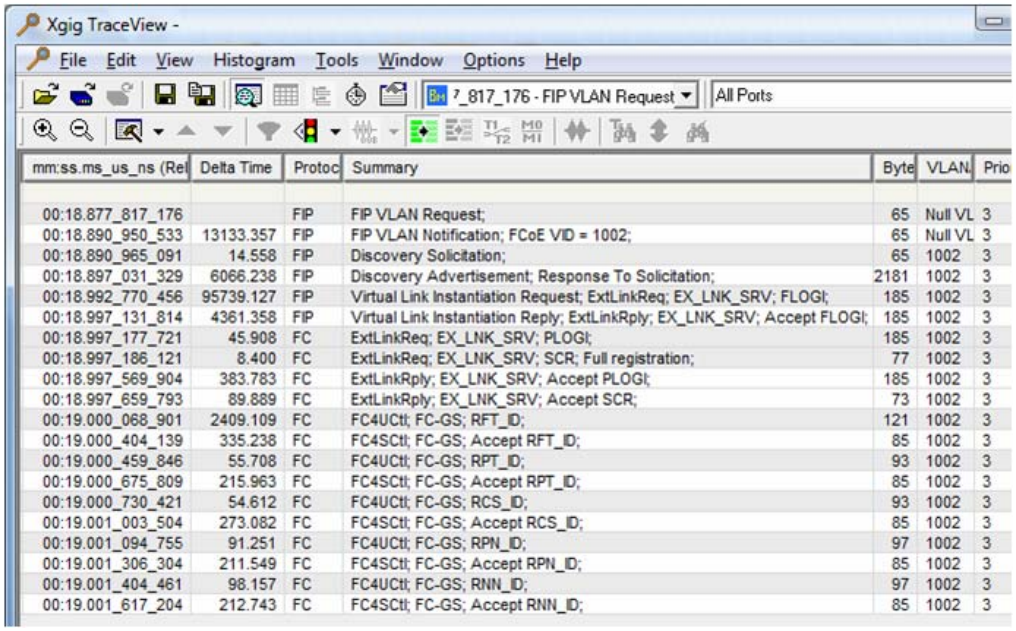
4. The switch responds with the FIP Discovery Advertisement/Response To Solicitation message. This message is enlarged to the maximum FCoE frame size. The delivery of this message proves that the network is capable of delivering maximum-sized FCoE frames. The message contains the Available For Login (A) bit, which must be ON so that end devices can log in. It also contains the fabric-provided MAC address/server-provided MAC address (SPMA/FPMA) setting, which specifies whether FCoE MAC addresses are provided by the switch (FPMA), or by a third-party server (SPMA). FPMA is the common way of providing MAC addresses.
5. The end devices attempt to log in to the switch with the FIP FLOGI message.
6. The switch responds with an Accept FLOGI or a Reject. The Accept FLOGI contains all the Fibre Channel characteristics. It also contains the new MAC address to be used by the end device for every FCoE communication. That MAC address typically starts with 0E:FC:00, followed by the 3-byte Fibre Channel source ID (S_ID).
7. When the login is successful, the regular FCoE frames flow. FCoE PLOGI and other regular Fibre Channel and FCoE initializations follow the FIP FLOGI.

Validation Process

1. Set up the Analyzer's TraceControl to capture FIP frames over the Ethernet VLAN by setting these frame types as trigger conditions. The FIP EtherType is 0x8914.
2. Configure each device (adapter, switch port, FCoE storage) to enable FCoE/FIP functions.
3. Capture a trace.
4. Use Analyzer Expert to examine the trace for errors on the FIP or FCoE protocol.
5. Open the trace in TraceView to view the FIP frames.

Validation Results

Figure 5-6 shows the FIP test results between the QLogic QLE8152 adapter and the Cisco Nexus 5000 FCoE switch.



mm:ss.ms_us_ns (Rel	Delta Time	Protoc	Summary	Byte	VLAN	Prio
00:18.877_817_176		FIP	FIP VLAN Request;	65	Null VL	3
00:18.890_950_533	13133.357	FIP	FIP VLAN Notification; FCoE VID = 1002;	65	Null VL	3
00:18.890_965_091	14.558	FIP	Discovery Solicitation;	65	1002	3
00:18.897_031_329	6066.238	FIP	Discovery Advertisement; Response To Solicitation;	2181	1002	3
00:18.992_770_456	95739.127	FIP	Virtual Link Instantiation Request; ExtLinkReq; EX_LNK_SRV; FLOGI;	185	1002	3
00:18.997_131_814	4361.358	FIP	Virtual Link Instantiation Reply; ExtLinkRply; EX_LNK_SRV; Accept FLOGI;	185	1002	3
00:18.997_177_721	45.908	FC	ExtLinkReq; EX_LNK_SRV; PLOGI;	185	1002	3
00:18.997_186_121	8.400	FC	ExtLinkReq; EX_LNK_SRV; SCR; Full registration;	77	1002	3
00:18.997_569_904	383.783	FC	ExtLinkRply; EX_LNK_SRV; Accept PLOGI;	185	1002	3
00:18.997_659_793	89.889	FC	ExtLinkRply; EX_LNK_SRV; Accept SCR;	73	1002	3
00:19.000_068_901	2409.109	FC	FC4UCtl; FC-GS; RFT_ID;	121	1002	3
00:19.000_404_139	335.238	FC	FC4SCtl; FC-GS; Accept RFT_ID;	85	1002	3
00:19.000_459_846	55.708	FC	FC4UCtl; FC-GS; RPT_ID;	93	1002	3
00:19.000_675_809	215.963	FC	FC4SCtl; FC-GS; Accept RPT_ID;	85	1002	3
00:19.000_730_421	54.612	FC	FC4UCtl; FC-GS; RCS_ID;	93	1002	3
00:19.001_003_504	273.082	FC	FC4SCtl; FC-GS; Accept RCS_ID;	85	1002	3
00:19.001_094_755	91.251	FC	FC4UCtl; FC-GS; RPN_ID;	97	1002	3
00:19.001_306_304	211.549	FC	FC4SCtl; FC-GS; Accept RPN_ID;	85	1002	3
00:19.001_404_461	98.157	FC	FC4UCtl; FC-GS; RNN_ID;	97	1002	3
00:19.001_617_204	212.743	FC	FC4SCtl; FC-GS; Accept RNN_ID;	85	1002	3

Figure 5-6. FIP Verification Results—Converged Network Adapter and FCoE Switch

From the captured trace, you can verify the following:

- FIP version used by both devices is 1.
- FCoE VLAN ID returned by the switch is 1002.
- The link is capable of delivering 2181-byte frames.
- The end device was able to log in to the switch with the FIP FLOGI.
- The switch sent an Accept FLOGI that specifies 01181C as the Fibre Channel Source ID and 0E:FC:00:01:18:1C as the MAC address.
- All subsequent FCoE messages use the source MAC address 0E:FC:00:01:18:1C for the end device.

In summary, the FIP link initialization process is successful—the end device successfully connects to the switch, and the FCoE virtual link is set up. Similar results should be seen on the FCoE target side. When the FCoE link is up, the end device sends an FIP Keep Alive frame every eight seconds so that the switch knows that the device is present even when the link is idle. If the switch does not receive FIP Keep Alives from a logged-in device, it will automatically disconnect that device with a FIP Clear Virtual Link message after a period of approximately 2.5 times the value of the variable FKA_ADV_PERIOD.

To validate that the switch disconnects end devices after 20 seconds, use the JDSU Jammer to remove all FIP Keep Alive messages from the link after an end device has successfully logged-in. The switch sends the FIP Clear Virtual Link message after 20 seconds.

Figure 5-7 shows that the switch requested that the end device send FIP Keep Alive frames every 8000 ms. The information is shown in the FIP Discovery Advertisement frames (both to All_ENode_MACs and All_FCF_MACs).

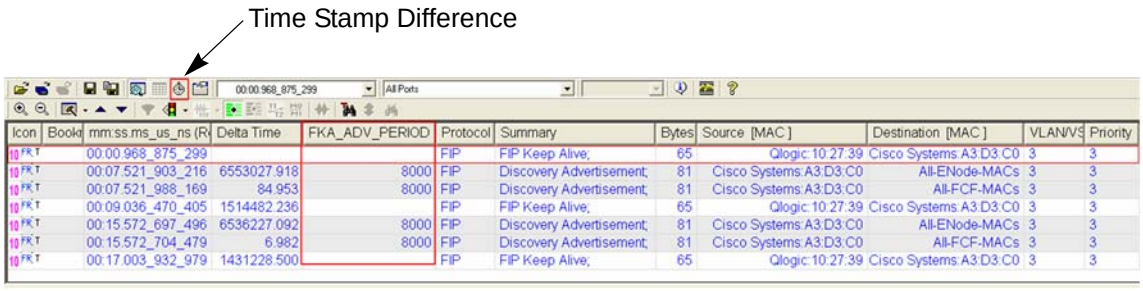

Time Stamp Difference
Icon
Book
mm:ss.ms_us_ns (R
Delta Time
FKA_ADV_PERIOD
Protocol
Summary
Bytes
Source [MAC]
Destination [MAC]
VLAN/VS
Priority
00:00.968_875_299 FIP FIP Keep Alive; 65 Qlogic:10:27:39 Cisco Systems:A3:D3:C0 3 3
00:07.521_903_216 6553027.918 8000 FIP Discovery Advertisement; 81 Cisco Systems:A3:D3:C0 All-ENode-MACs 3 3
00:07.521_988_169 84.953 8000 FIP Discovery Advertisement; 81 Cisco Systems:A3:D3:C0 All-FCF-MACs 3 3
00:09.036_470_405 1514482.236 FIP FIP Keep Alive; 65 Qlogic:10:27:39 Cisco Systems:A3:D3:C0 3 3
00:15.572_697_496 6536227.092 8000 FIP Discovery Advertisement; 81 Cisco Systems:A3:D3:C0 All-ENode-MACs 3 3
00:15.572_704_479 6.982 8000 FIP Discovery Advertisement; 81 Cisco Systems:A3:D3:C0 All-FCF-MACs 3 3
00:17.003_932_979 1431228.500 FIP FIP Keep Alive; 65 Qlogic:10:27:39 Cisco Systems:A3:D3:C0 3 3

Figure 5-7. Test Results of FIP Keep_Alive and Discovery Advertisement for FCoE Virtual Link Maintenance

Click in the icon column (Figure 5-7) to calculate the time stamp difference between the two adjacent FIP Keep Alive frames. Figure 5-8 shows a difference of 8.07 seconds.

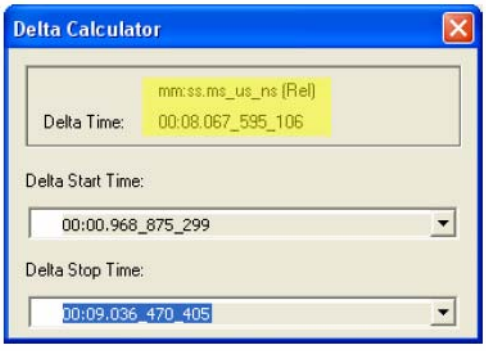


Figure 5-8. Time Difference Between Two Adjacent FIP Keep_Alive Frames

Unlike end devices, the FCoE switches do not send FIP Keep Alive messages. Instead, FCoE switches send FIP Discovery Advertisement messages every eight seconds, which notifies end devices that the switches are still present. Figure 5-9 shows the time difference between two FIP Discovery Advertisements. Both the E_Node and switch maintain the FCoE virtual link by sending FIP Keep Alive messages every eight seconds and FIP Discovery Advertisements every eight seconds.

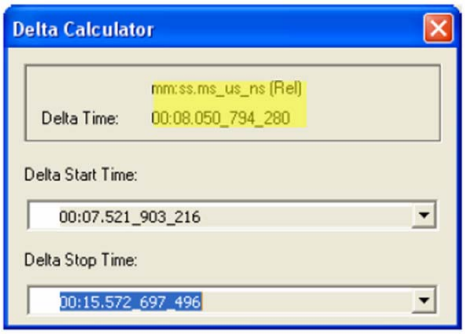


Figure 5-9. Time Difference Between Two Adjacent FIP Discovery Advertisement Multicast Frames (to All_E_Nodes)

Figure 5-10 shows a capture where the Jammer has removed all FIP Keep Alive messages from the link after Keep Alive 3 message was sent. The Cisco switch cleared the virtual link with a FIP Clear Virtual Link message 20 seconds after the last FIP Keep Alive message. After that, the adapter logs out and initiates a new login with the FIP FLOGI message.

	00:32.278_755_171	GE Switch(1,1,2)	1 - ExtLinkRply	FC	ExtLinkRply; EX_LNK_SRV; Accept PRLI; Request executed;
Keep Alive 2	00:37.380_826_966	GE Server-JM(1,1,1)	1 - Ether Frame	FIP	FIP Keep Alive;
	00:39.839_337_256	GE Switch(1,1,2)	1 - Ether Frame	FIP	Discovery Advertisement;
	00:39.839_343_129	GE Switch(1,1,2)	1 - Ether Frame	FIP	Discovery Advertisement;
Keep Alive 3	00:45.447_737_321	GE Server-JM(1,1,1)	1 - Ether Frame	FIP	FIP Keep Alive;
	00:47.929_682_611	GE Switch(1,1,2)	1 - Ether Frame	FIP	Discovery Advertisement;
	00:47.929_759_434	GE Switch(1,1,2)	1 - Ether Frame	FIP	Discovery Advertisement;
	00:55.939_976_191	GE Switch(1,1,2)	1 - Ether Frame	FIP	Discovery Advertisement;
	00:55.940_056_979	GE Switch(1,1,2)	1 - Ether Frame	FIP	Discovery Advertisement;
	01:03.980_460_979	GE Switch(1,1,2)	1 - Ether Frame	FIP	Discovery Advertisement;
	01:03.980_467_586	GE Switch(1,1,2)	1 - Ether Frame	FIP	Discovery Advertisement;
SW Clean Virtual	01:05.451_984_839	GE Switch(1,1,2)	1 - Ether Frame	FIP	FIP Clear Virtual Link;
	01:12.050_729_459	GE Switch(1,1,2)	1 - Ether Frame	FIP	Discovery Advertisement;
	01:12.050_809_666	GE Switch(1,1,2)	1 - Ether Frame	FIP	Discovery Advertisement;
	01:20.150_995_061	GE Switch(1,1,2)	1 - Ether Frame	FIP	Discovery Advertisement;
	01:20.151_075_639	GE Switch(1,1,2)	1 - Ether Frame	FIP	Discovery Advertisement;
	01:28.211_470_011	GE Switch(1,1,2)	1 - Ether Frame	FIP	Discovery Advertisement;
	01:28.211_550_187	GE Switch(1,1,2)	1 - Ether Frame	FIP	Discovery Advertisement;
FLOGI	01:33.655_994_466	GE Server-JM(1,1,1)	1 - ExtLinkReq	FIP	Virtual Link Instantiation Request; ExtLinkReq; EX_LNK_SRV; FLOGI;
	01:33.684_649_101	GE Switch(1,1,2)	1 - ExtLinkReq	FC	ExtLinkReq; EX_LNK_SRV; PLOGI;
	01:33.684_818_226	GE Switch(1,1,2)	1 - ExtLinkRply	FIP	Virtual Link Instantiation Reply; ExtLinkRply; EX_LNK_SRV; Accept FLO
	01:33.684_920_386	GE Server-JM(1,1,1)	1 - ExtLinkReq	FC	ExtLinkReq; EX_LNK_SRV; PLOGI;

Figure 5-10. Link De-instantiated by the Switch 20 Seconds after the Last Keep Alive Frame

Figure 5-11 shows the time difference between the last FIP Keep Alive Sent message and the FIP Clear Virtual Link message from the switch.

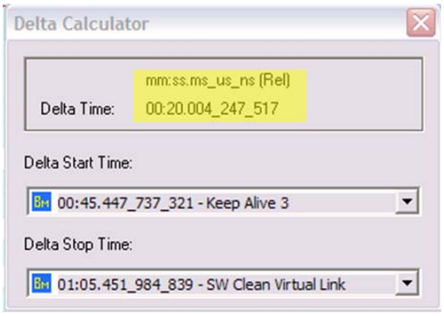


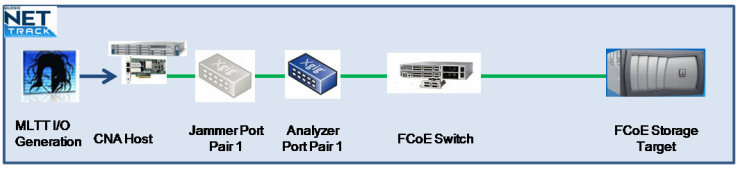
Figure 5-11. Time Difference Between the Last FIP Keep Alive Sent and the FIP Clear Virtual Link Message from the Switch

Validation Step 3—FCoE Function and I/O Tests

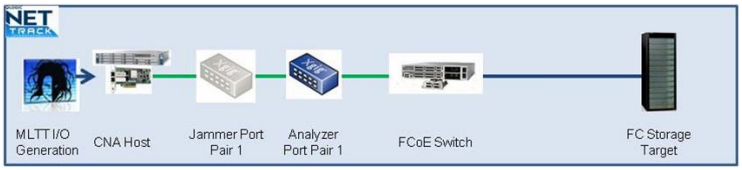
Figure 5-12 shows the three test configurations that validate FCoE function and I/O in the FCoE fabric:

- Converged Network Adapter host > FCoE switch > FCoE storage
- Converged Network Adapter host > FCoE switch > direct-attached Fibre Channel storage
- Converged Network Adapter host > FCoE switch > Fibre Channel switch > Fibre Channel storage

Converged Network Adapter Host > FCoE Switch > FCoE Storage



Converged Network Adapter Host > FCoE Switch > Direct Attached Fibre Channel Storage



Converged Network Adapter Host > Fibre Channel Switch > Fibre Channel Storage

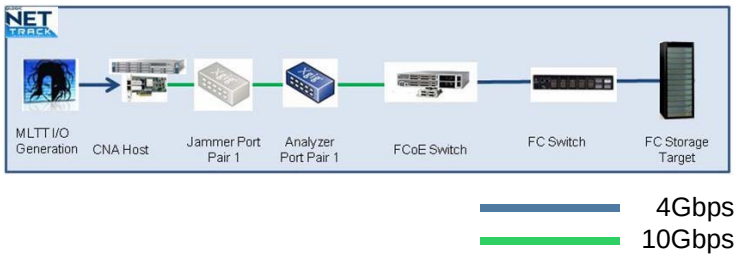


Figure 5-12. I/O Test with Fibre Channel SAN Storage Setup

Validation Process

The following validation steps are used for all three test configurations:

1. Install the Medusa test tools (MLTT) on the Converged Network Adapter host.
2. Map the target device from the MLTT client interface.

3. Configure the test as follows:
   - MLTT test pattern: mixed read/write operations, 100 percent read and write traffic
   - Data size: 512K
   - Queue depth: eight
4. Use MLTT to calculate IOPS and latency for the three test configurations, and verify data integrity.
5. Set up the Analyzer's TraceControl to capture FCoE frames over the Ethernet VLAN by setting these frame types as trigger conditions. The FCoE EtherType is 0x8906.
6. Use Expert to analyze the trace. Compare the details of the I/O performance for the three test configurations, and identify any errors in the FCoE traffic.

Validation Results

Figure 5-13 shows the MLTT I/O performance test for the three network options.

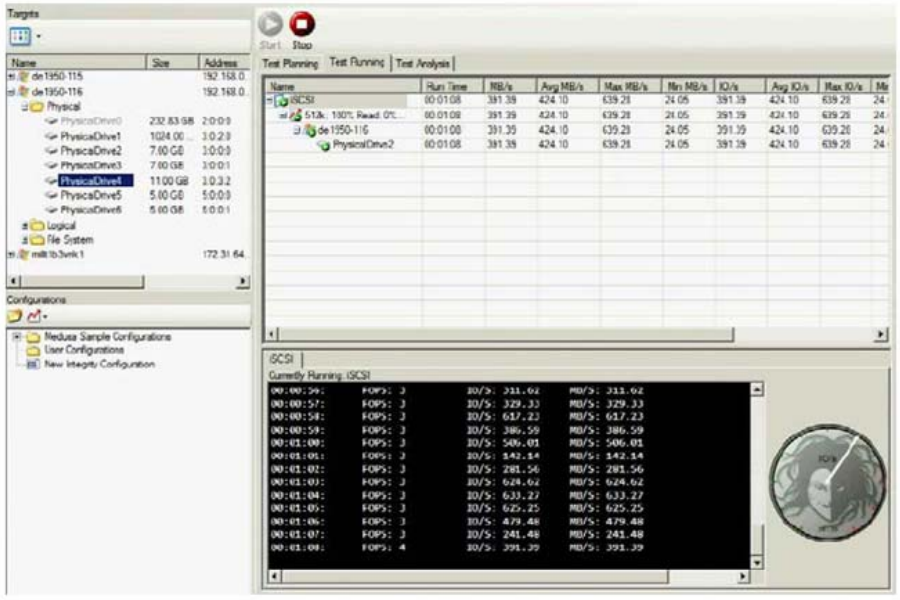


Figure 5-13. MLTT Configures Various I/O Applications to Verify I/O Benchmarking Performance of Different Storage Networks

Similar to previous test results, the captured FCoE traces listed in Table 5-1 show no FCoE frame errors.

Table 5-1. I/O Performance Comparison Among FCoE Storage Network Scenarios

Topology Option	Throughput (MBps / Percent of Maximum Bandwidth)	SCSI I/O Statistics		Exchange Statistics			Pending Exchange	
		IOPS (IO/s)	Average Read Data Time	Average Completion Time (ms)	Completion Rate (per sec)	Byte/ Exchange	Pending I/O (per sec)	Oldest Pending Read (ms)
Direct Attach FCoE Storage	641 / 64%	1168	5.28	5.73	1168	524,288	8	0.39
Direct Attach Fibre Channel Storage	390 / 98%	703	1.28	5.62	704	524,288	1	9.37
Fibre Channel SAN Storage	390 / 98%	736	1.28	5.59	736	524,288	5	4.84

The results show that the direct attached Fibre Channel storage topology and Fibre Channel SAN storage topology share similar I/O performance. The throughput is nearly 100 percent of total available bandwidth. These results prove that the stateless FCoE overhead does not impact Fibre Channel SAN performance.

The direct FCoE storage topology has higher available bandwidth (10GbE link), so the throughput is higher at 641MBps. However, the bandwidth usage is only 64 percent of the total available bandwidth. The bandwidth may have been limited by the storage device in this case, whereas it was limited by the network in the other two cases. This conclusion is based on the Queue Depth and Data Size traffic parameters.

No data integrity errors were detected by the MLTT tool in the three tests.

Validation Step 4—Validate PFC and Lossless Link

The critical factor for a successful FCoE deployment is the ability of the virtual Ethernet link to carry lossless FCoE traffic. Figure 5-14 shows the test configuration to validate PFC function and to validate that the link is lossless.

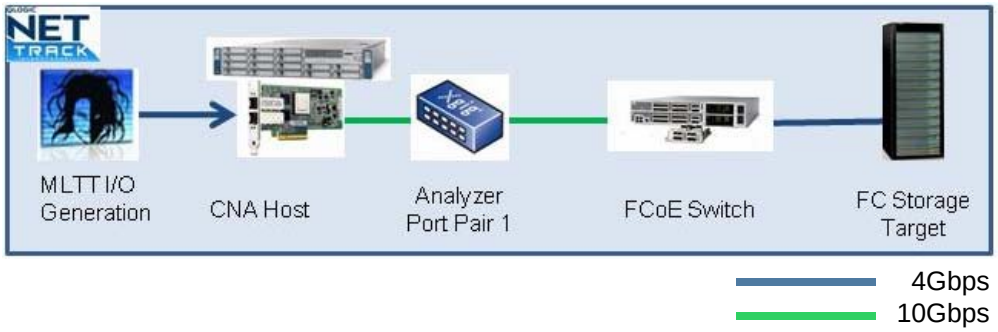


Figure 5-14. PFC Validation Setup

Validation Process

The test assumes the following:

- The hosts have storage LUNs mapped and can generate I/O to storage devices.
- The Analyzer connects to the FCoE switch with a 10 Gbps link, and the FCoE switch connects to the Fibre Channel storage target with a 4Gbps link.

1. Configure the MLTT test as follows:
   - Data size: 512K (writes only)
   - Queue depth: eight
   - Bandwidth: 100 percent at the Fibre Channel link between the storage and the FCoE switch
2. Launch the test.

   The test is designed to overwhelm the 4Gbps link with Write commands so that the Fibre Channel link runs out of buffer-to-buffer credits, and the FCoE switch issues PFC Pause frames to pause the Converged Network Adapter transmitter.

3. To validate that PFC Pause frames are sent to reduce the traffic, configure TraceControl to capture a large buffer when it encounters a PFC Pause frame. To set up a capture:

   a. Select the trigger mode.

   b. Set up the trigger condition to capture on a PFC Pause frame and 80 percent of post-fill. The capture memory is a circular buffer that stops capturing when 80 percent of the buffer is filled after the encountering the trigger condition.

If TraceControl does not encounter the trigger condition on a PFC frame, it may be because the Fibre Channel link is not saturated—increase the traffic rate until TraceControl stops capturing. If the traffic throughput on the 10GigE link reaches the theoretical maximum bandwidth, and TraceControl still does not encounter the trigger condition, it may be because PFC is not properly configured, and the link is not lossless. In this case, ensure that PFC is enabled, and restart the validation at Step 1.

When TraceControl stops capturing, it means that at least one PFC frame crossed the link. Open the capture in Expert so that it analyzes all pauses in the trace, and check for errors. Click **Report** in the Expert tool bar to generate a report showing the average and maximum pause times (Table 5-2).

Validation Results

Figure 5-15 shows the captured PFC frames. The Cisco FCoE switch pauses the FCoE traffic in the other direction using a PFC frame with the maximum pause time (3,355 µs) at priority 3. The switch releases the pause by sending a PFC frame with a zero pause time before the previous PFC frame expires. TraceView measures the effective pause time to be 29.66 µs. In this case, no FCoE frame in the other direction is being transmitted between the two PFC frames, which indicates that the pause mechanism on the FCoE link is working as expected, and is equivalent to the buffer-to-buffer credit mechanism on the Fibre Channel link.

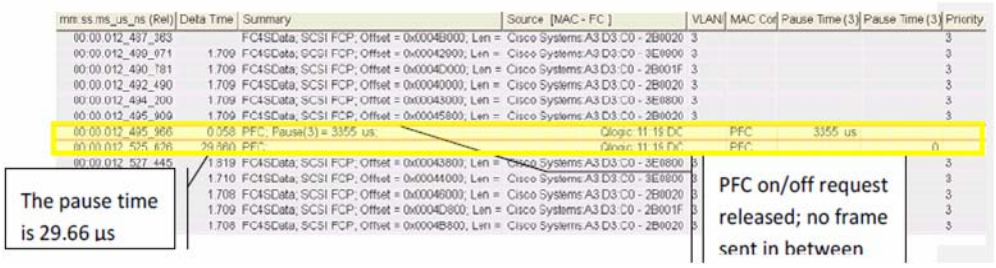


Figure 5-15. Trace Shows the PFC Event to Pause the Transmitting FCoE Data Frame Further

As described earlier, the Expert software analyzes all the pauses in a capture and reports critical errors, such as `Frame Received while PFC Class Paused` (Figure 5-16). These errors are serious—it is equivalent to sending a Fibre Channel frame on a link without waiting for a credit. In such a case, the switch will probably drop the frames, which will cause the Fibre Channel protocol to abort the sequence, and force the initiator to re-issue the entire sequence. If this is a recurring error, the link performance declines dramatically.

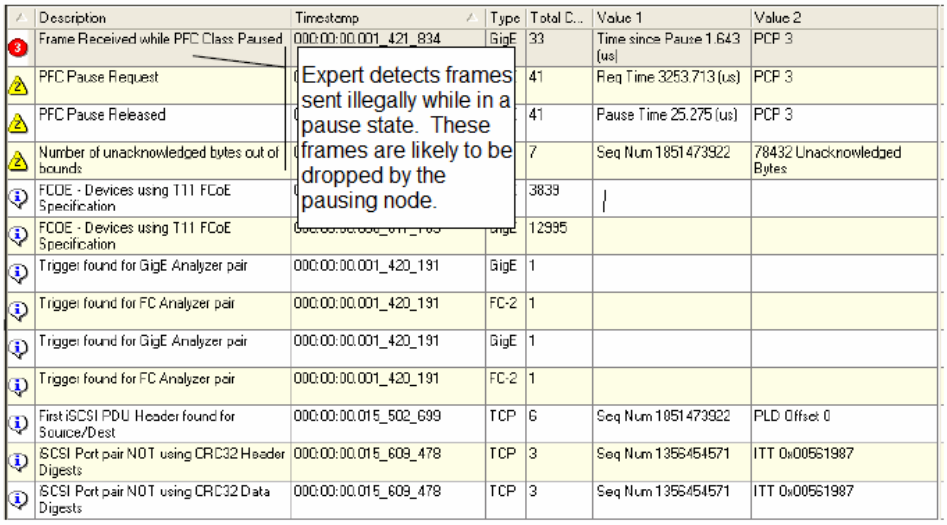


Figure 5-16. Expert Validating the Reaction to PFC Pause Frames

In Expert, click **Report** to create a detailed report on all protocol layers (Figure 5-17). This report gives you detailed information about network issues and statistics.

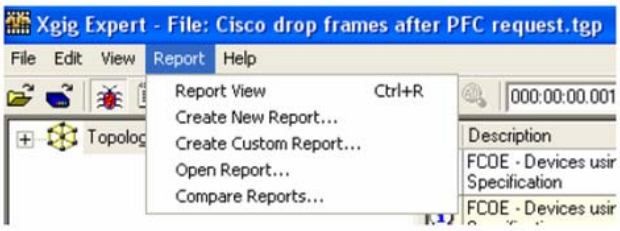


Figure 5-17. Expert Generates Reports with Detailed Protocol Analysis and Statistics

Table 5-2 shows some of the statistics reported by Expert software for evaluating PFC functions and performance. The sample data indicates that there 33 frames were received during PFC pauses, and as a result, the link may not guarantee lossless traffic for the FCoE application. Furthermore, the large variance between the maximum and minimum exchange completion time is caused by the PFC pause.

Table 5-2. FCoE Application Performance Statistics Evaluated by Expert

Throughput, any to any (MBps)	Average Exchange Completion Time (ms)	Maximum Exchange Completion time (ms)	Minimum Exchange Completion time (ms)	Average Pause Time (µs)	Maximum Pause Time (µs)	Frames Received while PFC Class Paused	PFC Request Count
384	4.6	7.4	1.3	28.35	38.31	33	41

Validation Step 5—ETS Verification

The purpose of this test is to verify that the switch can continuously adjust bandwidth with different traffic priorities. This test verifies that the unified fabric manages congestion between TCP, iSCSI, and FCoE traffic, and measures application performance.

Validation Process

The test assumes that the hosts have mapped storage device LUNs and can generate I/O to storage devices. The test should include two traffic classes: one for FCoE and one for combined traffic, including iSCSI and others. If you want to isolate and validate the switch by itself, you can also replace the Converged Network Adapter hosts and storage targets with the Load Tester, which acts as a traffic generator on both sides of the switch.

1. Configure ETS parameters so that:
   - FCoE traffic has its own priority.
   - PFC is enabled.
   - Bandwidth is assigned to the FCoE priority. You can also allow the FCoE switch to operate at its default value (50/50).

2. Run all the tests with the Write configuration (Figure 5-18), and then repeat all the tests with the Read configuration (Figure 5-19).

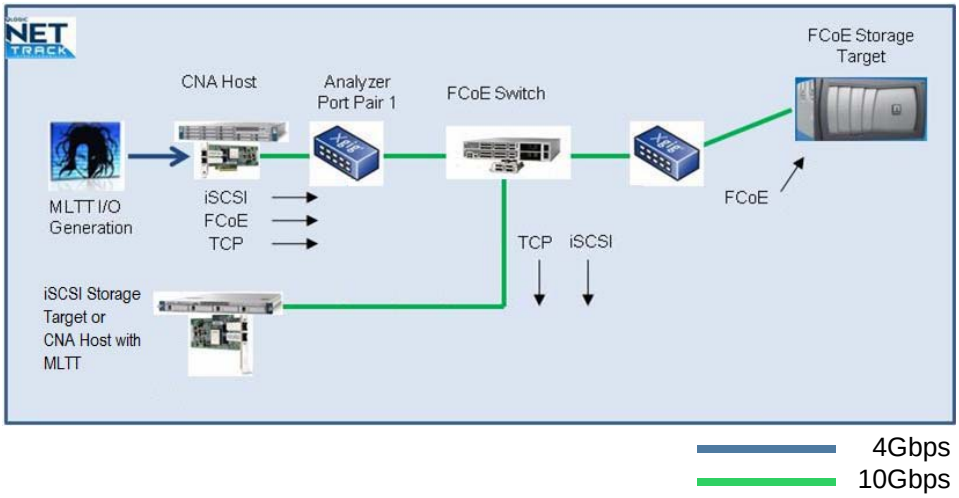


Figure 5-18. Verifying ETS with Write Operations to the Storage

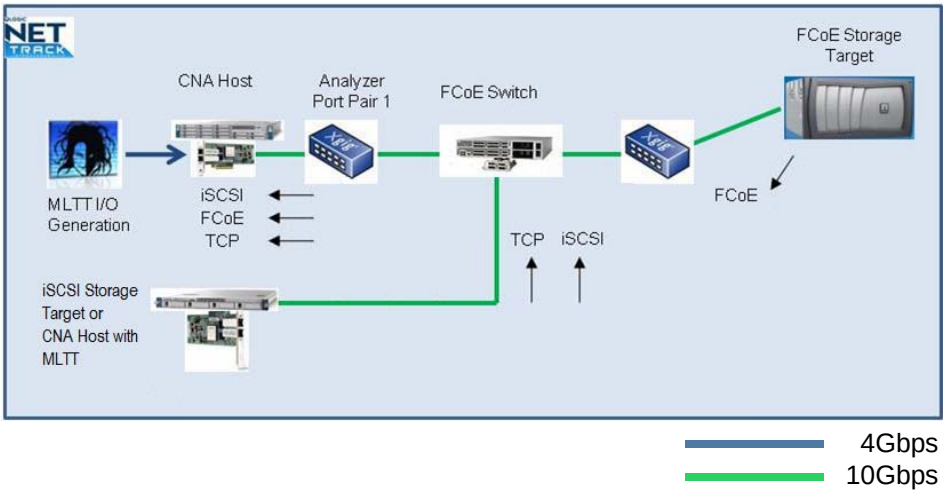


Figure 5-19. Verifying ETS with Read Operations from the Storage

3. Start the FCoE and other traffic (TCP, iSCSI) simultaneously in the MLTT application to saturate the link.
4. Use TraceControl to monitor link throughput.
5. Pause I/O from each application in a round-robin manner, and verify that the free bandwidth is being distributed to the other traffic classes. Figures 5-21, 5-22, and 5-25 show the results of pausing I/O in a round-robin manner.
6. Capture a trace.

7. Use Expert to analyze the capture, and verify the following:
   - No error occurs on FCoE traffic
   - Link throughput per priority in MBps (Figure 5-20)
   - Each traffic class can still generate I/O at its ETS setting when the link is saturated.
   - Exchange completion time

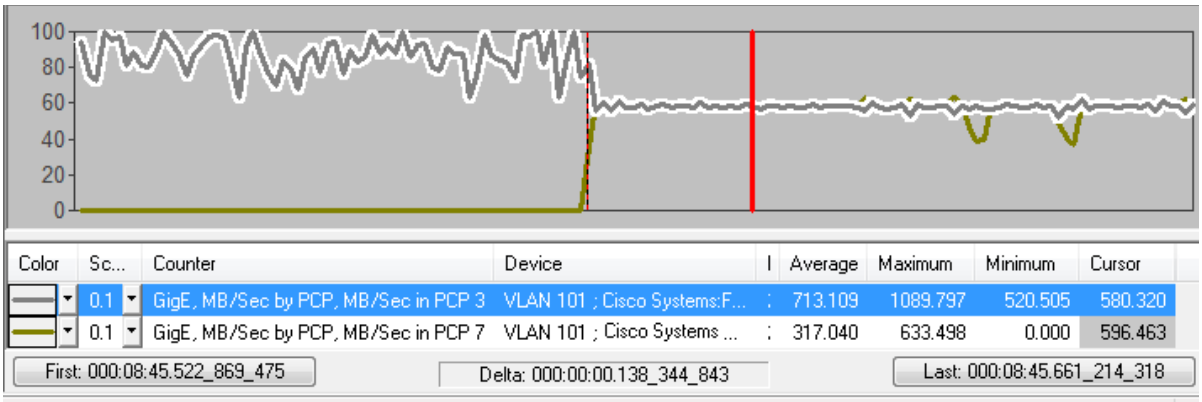


Figure 5-20. Throughput per Priority

Validation Results

Table 5-3 lists the port settings to verify the ETS setup. ETS allocates 50 percent of the bandwidth to FCoE and 50 percent of the bandwidth to other traffic. The other traffic is regular TCP traffic.

Table 5-3. Analyzer Port Setup Reference for Monitoring Links

Ports	Direction	Description
Storage (1,1,1)	Storage > Switch	FCoE storage read data
Switch (1,1,2)	Switch > Storage	FCoE storage write data
Server (1,2,1)	Server > Switch	Server I/O out (write)
Switch (1,2,2)	Switch > Server	Server I/O in (read)

Figure 5-21 shows that the FCoE traffic is read at the rate of 837MBps. The throughput numbers on ports (1,1,1) and (1,2,2) are equal, indicating that all traffic from the FCoE storage is going to the server. The throughput of 837MBps represents the maximum rate for FCoE in this configuration.

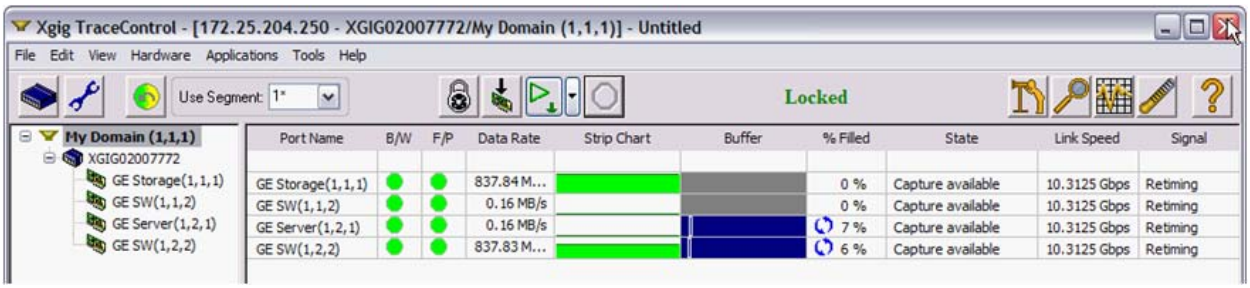

Xgig TraceControl - [172.25.204.250 - XGIG02007772/My Domain (1,1,1)] - Untitled
File Edit View Hardware Applications Tools Help
Use Segment: 1*
Locked
My Domain (1,1,1)
XGIG02007772
GE Storage(1,1,1)
GE SW(1,1,2)
GE Server(1,2,1)
GE SW(1,2,2)
Port Name B/W F/P Data Rate Strip Chart Buffer % Filled State Link Speed Signal
GE Storage(1,1,1) 837.84 M... 0 % Capture available 10.3125 Gbps Retiming
GE SW(1,1,2) 0.16 MB/s 0 % Capture available 10.3125 Gbps Retiming
GE Server(1,2,1) 0.16 MB/s 7 % Capture available 10.3125 Gbps Retiming
GE SW(1,2,2) 837.83 M... 6 % Capture available 10.3125 Gbps Retiming

Figure 5-21. FCoE Traffic in a Read Operation from Storage

In Figure 5-22, TCP read traffic starts while maintaining the FCoE read operations. The traffic throughput from the switch on GE SW(1,2,2) increases to approximately the full line rate (1176MBps). Because the traffic reached full line rate, the switch enforced ETS parameters. The FCoE traffic decreases to 600MBps, as seen on port GE Storage(1,1,1). The difference between the full line rate (1,176MBps) and the FCoE traffic (600MBps) is the TCP traffic (576MBps). The speed of 576MBps represents 49 percent of the full rate, which corresponds to the switch ETS configuration of 50 percent for FCoE and 50 percent for others. The switch is behaving as expected.

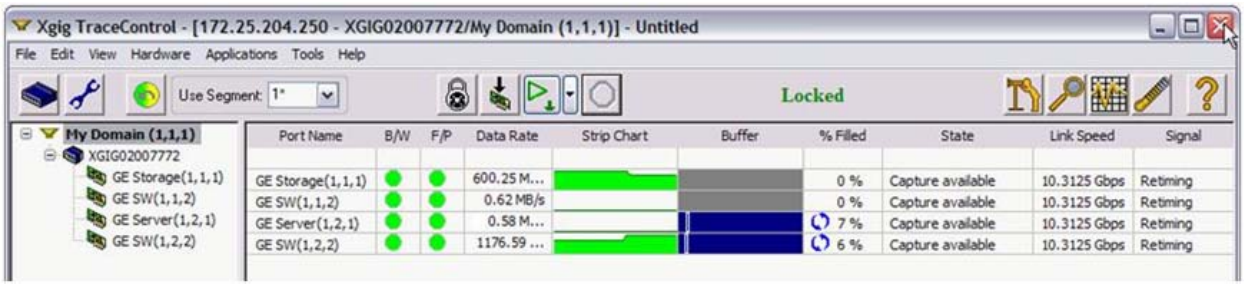

Xgig TraceControl - [172.25.204.250 - XGIG02007772/My Domain (1,1,1)] - Untitled
File Edit View Hardware Applications Tools Help
Use Segment: 1*
Locked
My Domain (1,1,1)
XGIG02007772
GE Storage(1,1,1)
GE SW(1,1,2)
GE Server(1,2,1)
GE SW(1,2,2)
Port Name B/W F/P Data Rate Strip Chart Buffer % Filled State Link Speed Signal
GE Storage(1,1,1) 600.25 M... 0 % Capture available 10.3125 Gbps Retiming
GE SW(1,1,2) 0.62 MB/s 0 % Capture available 10.3125 Gbps Retiming
GE Server(1,2,1) 0.58 M... 7 % Capture available 10.3125 Gbps Retiming
GE SW(1,2,2) 1176.59 ... 6 % Capture available 10.3125 Gbps Retiming

Figure 5-22. FCoE Traffic Plus TCP Traffic in a Read from Storage

To validate that the switch is not disrupting FCoE traffic while reducing it, capture a trace of the full line rate traffic, and then open it in Expert for analysis. Because the capture is taken on both sides of the switch, it is possible to run a cross-port analysis in Expert.

Figure 5-23 shows how to enable a cross-port analysis in Expert's Edit / Preferences dialog. In a cross-port analysis, Expert ensures that each frame is making its way across the switch, and reports frame losses and the latency through the switch. In addition, Expert reports aborted sequences and retransmitted FCoE, iSCSI, and TCP traffic. Expert reports all symptoms and network anomalies indicating frame losses and performance degradation.

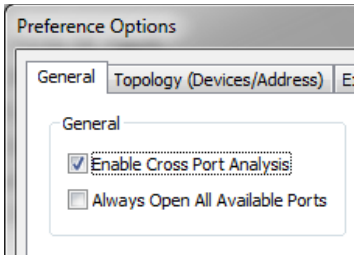


Figure 5-23. Enabling Cross Port Analysis in Expert

However, because the FCoE traffic has PFC enabled, the decrease in FCoE traffic is a result of the pause frames sent by the switch to the storage device. Figure 5-24 shows pause requests and pause releases sent only to the storage port.

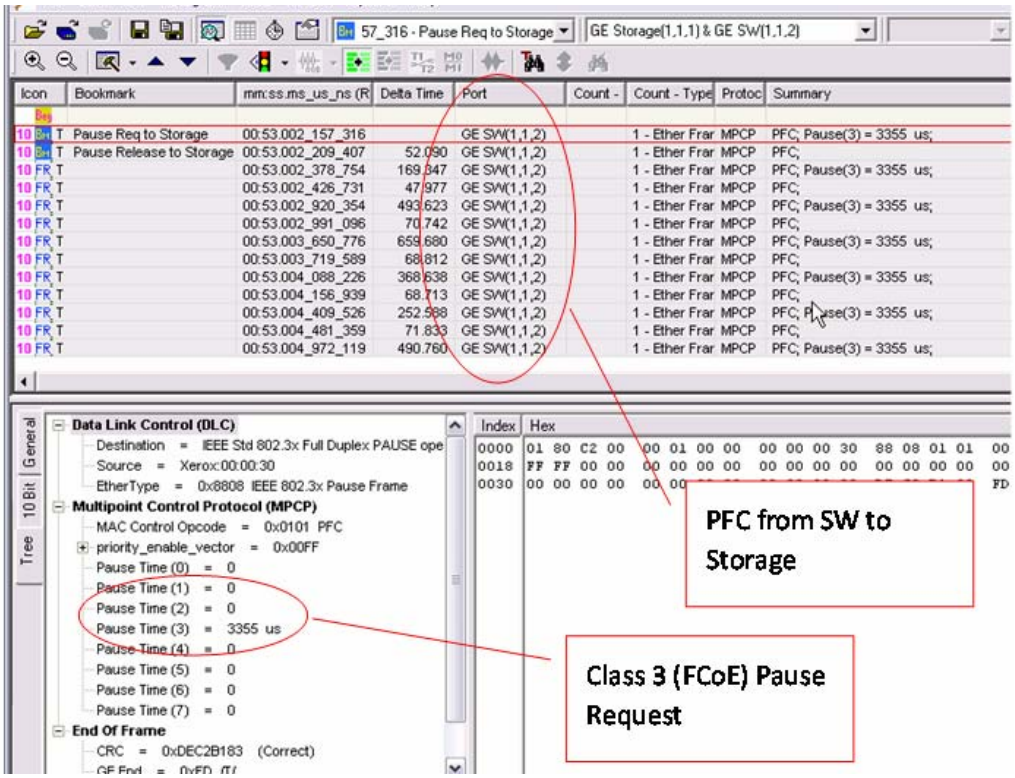

57_316 - Pause Req to Storage
GE Storage(1,1,1) & GE SW(1,1,2)
Pause Req to Storage
Pause Release to Storage
Data Link Control (DLC)
Multipoint Control Protocol (MPCP)
Pause Time (3) = 3355 us
PFC from SW to Storage
Class 3 (FCoE) Pause Request

Figure 5-24. Switch Issue Pause to Storage

The next step is to verify that the switch allows more TCP traffic when more bandwidth is available. Figure 5-25 shows that the TCP traffic increases to 741MBps after stopping the FCoE traffic. Again, the switch is behaving properly.

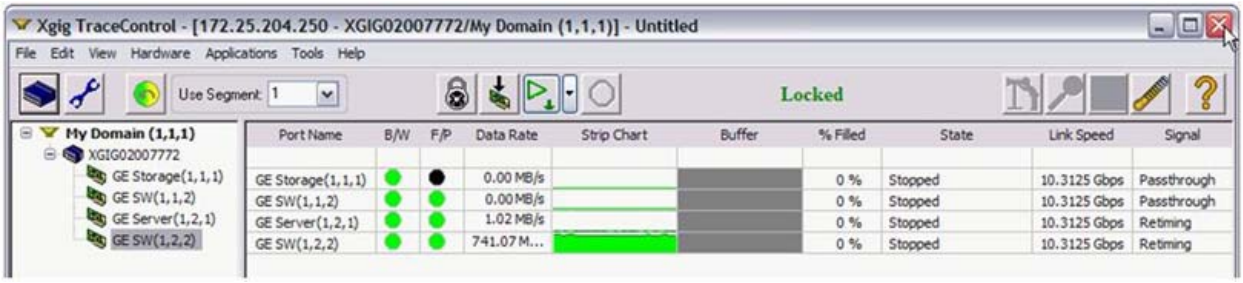

Xgig TraceControl - [172.25.204.250 - XGIG02007772/My Domain (1,1,1)] - Untitled
File Edit View Hardware Applications Tools Help
Use Segment: 1
Locked
My Domain (1,1,1)
XGIG02007772
GE Storage(1,1,1)
GE SW(1,1,2)
GE Server(1,2,1)
GE SW(1,2,2)
Port Name B/W F/P Data Rate Strip Chart Buffer % Filled State Link Speed Signal
GE Storage(1,1,1) 0.00 MB/s 0 % Stopped 10.3125 Gbps Passthrough
GE SW(1,1,2) 0.00 MB/s 0 % Stopped 10.3125 Gbps Passthrough
GE Server(1,2,1) 1.02 MB/s 0 % Stopped 10.3125 Gbps Retiming
GE SW(1,2,2) 741.07 M... 0 % Stopped 10.3125 Gbps Retiming

Figure 5-25. LAN Traffic Only

Table 5-4 summarizes the throughput achieved for the previous three tests.

Table 5-4. Summary of Throughput Characteristic

	Storage to Switch (MBps)	LAN Traffic (MBps)	Switch to Adapter (MBps)	Error
FCoE Traffic Only	837.84	0	837.83	No
FCoE Plus TCP	600.25	576.34	1176.59	No
LAN Traffic Only	0	741.07	741.07	No

The test continues cycling FCoE traffic and TCP traffic on and off. The results show that the switch continues to enforce the ETS configuration, allows both TCP and FCoE to use additional bandwidth when available, and maintains minimum guaranteed bandwidth to both traffic classes.

Figure 5-26 shows the relationship between FCoE, TCP, and pause frames. Before the start of TCP traffic, server ingress is equal to the FCoE traffic. When the TCP traffic starts, server ingress is the sum of the FCoE and TCP traffic. The switch starts sending PFC pause requests to the storage when the TCP traffic starts to slow the FCoE traffic going to the server.

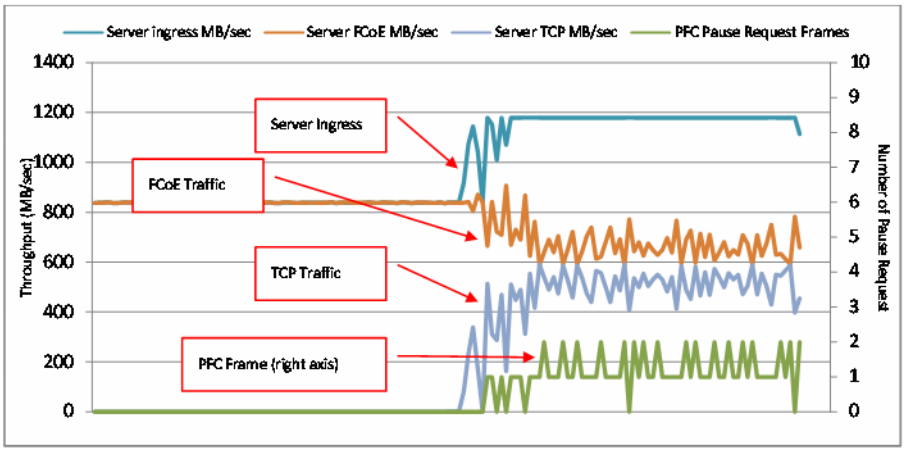


Figure 5-26. Relationship Between FCoE, TCP Traffic, and Pause Frames

Exchange Completion Time (Latency Monitor)

Latency is the measurement of how fast commands complete. The time between command issue and status received is measured in milliseconds (ms). Figure 5-27 shows the Fibre Channel read-response time in relationship to changes in traffic load. The graph also shows increases in maximum and average response times between the switch and the storage, while the switch issues pause requests to the storage device. The read latency increases from about 10ms to 12.5ms.

Figure 5-27. Read Latency Measurement

In this test, the link congestion did not happen at the host with the oversubscription scenario. Instead, the switch was congested. This is because of the multiplexing oversubscription at the egress port. The two ingress links sent a total of over 100 percent of traffic bandwidth to the egress port, causing the congestion and generation of PFC pause requests through the FCoE ingress port to the FCoE target. As Validation Step 7—Virtualization Verification will show, to congest the converged network adapter initiator, you can leverage the virtual machine carrying various applications and unified storage to avoid traffic multiplexing at the switch.

Validation Step 6—iSCSI Function and I/O Test

This test verifies that FCoE and iSCSI (IP) traffic is functioning well within the unified fabric environment. The key measurement parameters are throughput, IOPS, and completed I/O exchange rate for each traffic type. Figure 5-28 shows the configuration for testing iSCSI and I/O functions.

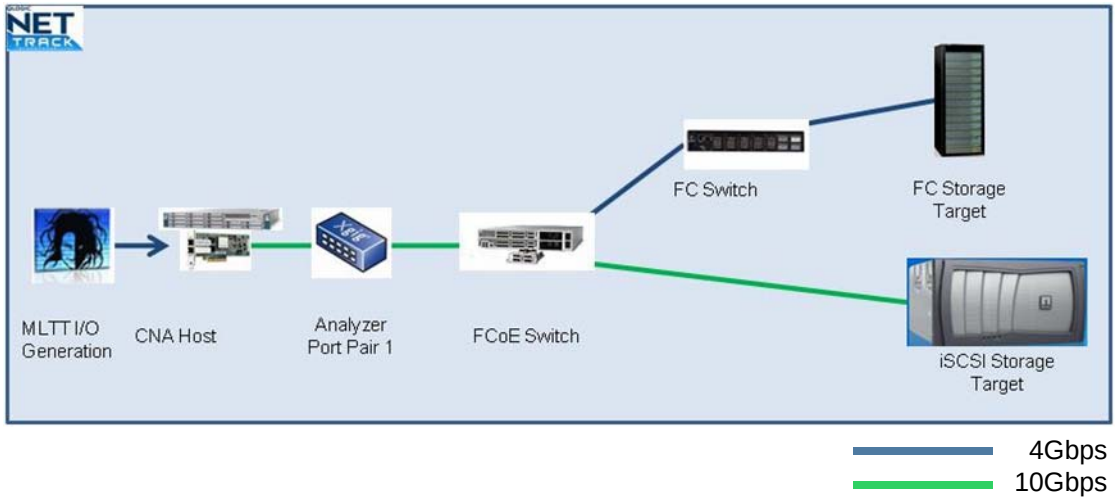


Figure 5-28. iSCSI Traffic Performance Validation Setup

Validation Process

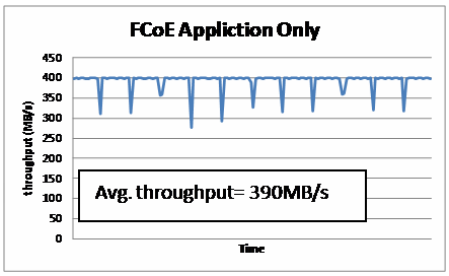
1. Enable the Converged Network Adapter to process FCoE and iSCSI data.
2. Configure the Analyzer to capture *any traffic*.
3. Use MLTT to map QLogic Fibre Channel storage devices.
4. Configure the test as follows:
   - Traffic: 100 percent read
   - Data size: 512K
   - Queue depth: eight.
5. Capture a trace for the Fibre Channel storage.
6. Use Expert to analyze the Fibre Channel storage trace, and calculate the average, minimum, and maximum throughput.
7. Map iSCSI storage devices.
8. Configure the test as follows:
   - Traffic: 100 percent read
   - Data size: 512K

- Queue depth: eight.

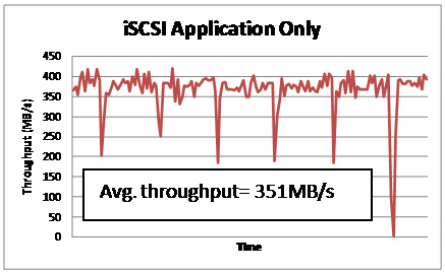
9. Capture a trace for the iSCSI storage.
10. Use Expert to analyze the iSCSI storage trace and calculate the average, minimum, and maximum throughput.
11. Capture a trace of the same read application for Fibre Channel and iSCSI storage together through the unified fabric.
12. Use Expert to analyze the combined Fibre Channel and iSCSI storage trace and calculate the average, minimum, and maximum throughput results per priority.
13. Compare performance results for FCoE storage, iSCSI storage, and converged FCoE and iSCSI storage.

Validation Results

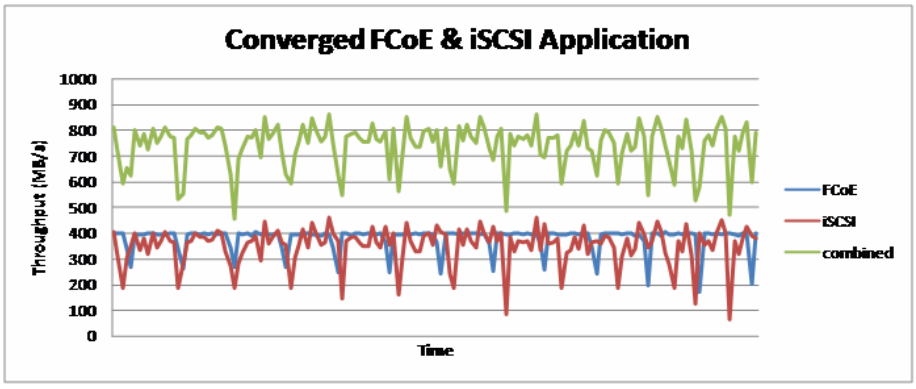
Figure 5-29 shows the throughput for each separate and combined application.



Adapter 100 percent read from only Fibre Channel storage



Adapter 100 percent read from only iSCSI storage



Adapter 100 percent read from Fibre Channel and iSCSI storage (at the same time) through the unified fabric

Figure 5-29. Throughput Results Showing the Fibre Channel and iSCSI Application Performance

Table 5-5 lists the average, minimum, and maximum data.

Table 5-5. Performance Comparison Between Separate and Combined Applications

Test option	FCoE only (MBps)	iSCSI Only (MBps)	FCoE and iSCSI (MBps)		
			FCoE	iSCSI	Combined
Average	390.670	365.706	380.240	351.356	740.507
Maximum	397.937	420.04	405.570	460.579	862.970
Minimum	275.835	0	169.261	63.266	454.195

The results show that:

- Application performance does not decline when running concurrently on the unified fabric.
- Converged throughput is the sum of the individual application performances. The 10G bandwidth is more efficient in a unified fabric.
- There is no congestion in the combined traffic option, because the storage device is the source of the bandwidth bottleneck, and not the converged link. The highest speed for Fibre Channel storage is 4Gbps, and 10Gbps for iSCSI storage. Fibre Channel storage throughput is about 90 percent of wire speed, and iSCSI storage throughput is about 35 percent of wire speed.
- There were no data integrity errors during the test.

In summary, this test clearly demonstrates the concept of a unified fabric. The unified fabric can converge SAN and LAN applications with no loss of performance or concerns about application interference. The unified fabric maximizes the bandwidth usage and lowers TCO by simplifying the network infrastructure and reducing the number of network devices.

Validation Step 7—Virtualization Verification

This validation step compares the performance of virtual machines on a single host with performance on multiple hosts. The test runs one iSCSI initiator in one virtual machine, and one FCoE initiator in another virtual machine as shown in Figure 5-30.

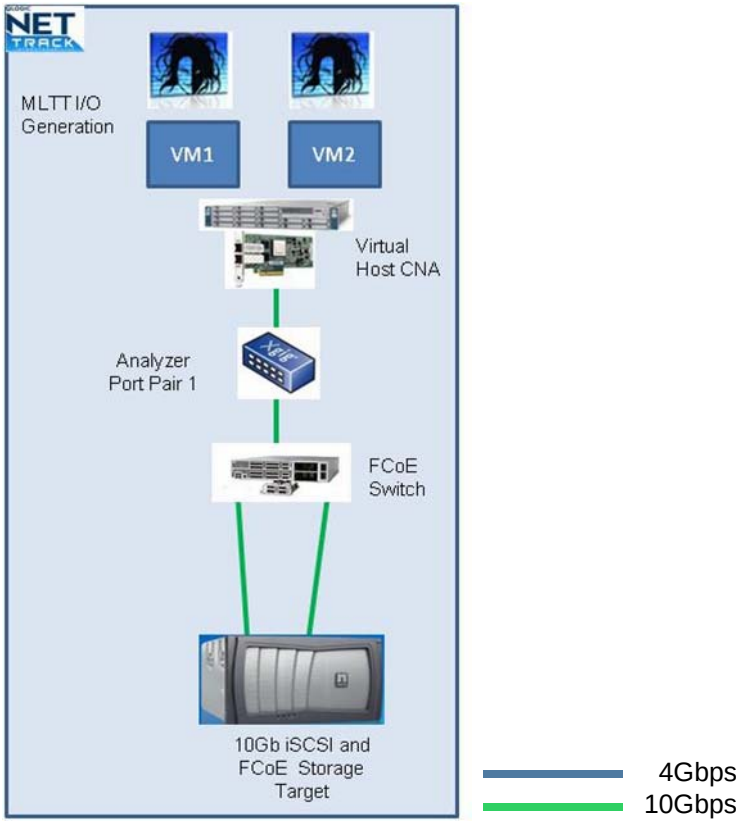


Figure 5-30. Setup for Verifying Virtualization

Validation Process

1. Set up two virtual machines on one server, and install MLTT on each virtual machine.
2. Set up ETS.
3. Configure the Analyzer to capture any frame.
4. Use MLTT to set up remote control from one virtual machine, so that you can control both virtual machines from one MLTT interface.
5. Use MLTT to map iSCSI storage and Fibre Channel storage devices for individual virtual machines.

6. Configure the test as follows:
   - Trigger: any PFC frame
   - Post fill: 90 percent after the trigger
   - Data size: 512K
   - Traffic: 100 percent read on both virtual machines
7. Use Analyzer to capture a trace.
8. Use Expert to analyze the trace, and calculate the average, minimum, and maximum throughput results per priority.
9. Compare I/O performance for the two virtual machines.

Repeat the entire test, using separate hosts for the iSCSI initiator and the FCoE initiator, and compare I/O performance.

Validation Results

One virtual machine ran the FCoE read application, and the other virtual machine ran the iSCSI read application. Figure 5-31 shows the throughput results for individual and combined applications.

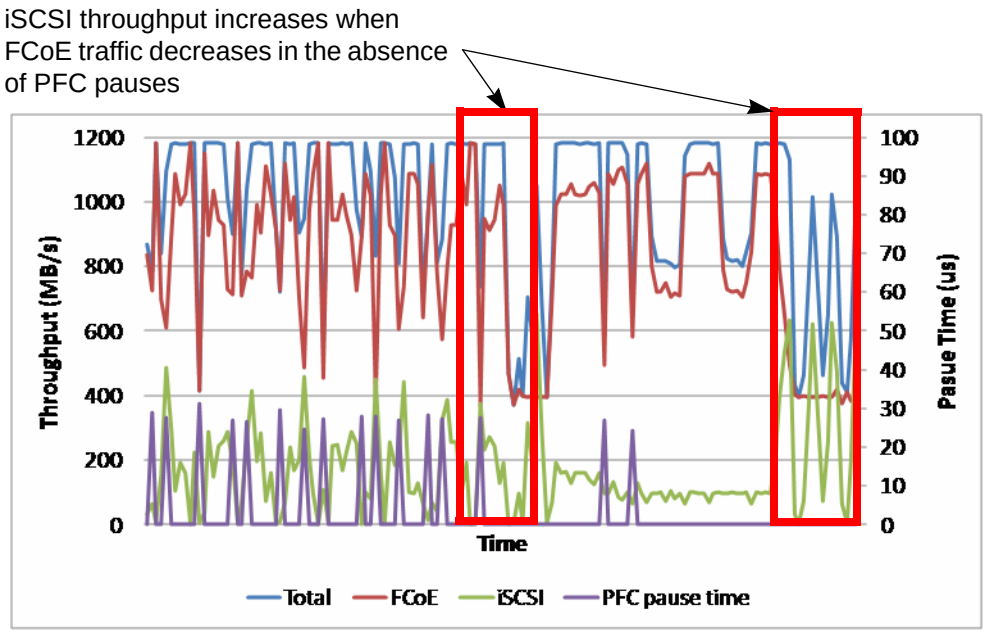


Figure 5-31. FCoE and iSCSI Application Throughput in a Virtual Environment

Table 5-6 lists the average minimum and maximum throughput values.

Table 5-6. FCoE and iSCSI Application Throughput when Running Separate Virtual Machines on One Physical Server

Test Option	FCoE (MBps)	iSCSI (MBps)	Combined (MBps)
Average	835.819	166.289	999.367
Maximum	1182.395	653.632	1182.395
Minimum	367.888	0	371.988

Figure 5-32 shows that the FCoE initiator issued 16 PFC pause requests, indicating congestion on the converged host side (with VMware in place).

Description	Timestamp	Source	Destination	Ports	Type	Total C...	Value 1	Value 2
Frame Received while paused during valid overlap	000:00:00.000_519_408	VLAN 3 ; Cisco Systems:A3:D3:C0	VLAN 3 ; 0e:fc:00:2b:00:23	2,1,1 / 2,1,2	GigE	10	Time since Pause 0.127 (us)	PCP 3
% of Sequence Burst Efficiency out of bounds	000:00:00.000_024_744	2b0100	2b0023	2,1,1 / 2,1,2	FC-4	139	Sequence ID 0xb8	23.75% Efficiency
PFC Pause Request	000:00:00.000_046_981	Qlogic:11:C3:80	IEEE Std 802.3x Full Duplex PAUSE operation	2,1,1 / 2,1,2	GigE	16	Req Time 3253.713 (us)	PCP 3
PFC Pause Released	000:00:00.000_075_896	Qlogic:11:C3:80	IEEE Std 802.3x Full Duplex PAUSE operation	2,1,1 / 2,1,2	GigE	16	Pause Time 28.915 (us)	PCP 3
Number of Unacknowledged Command PDUS out of bounds	000:00:00.001_860_955	192.168.0.8 ; e77c	192.168.0.21 ; 0cbc (iSCSI) ; 0001	2,1,1 / 2,1,2	iSCSI	28	ITT 0xfcbd1000	16 Unacknowledged Cmds
IP TTL at 01 - almost expired	000:00:00.000_064_076	192.168.0.8	192.168.0.21	2,1,1 / 2,1,2	IP	245	Protocol Type 6	
FCOE - Devices using T11 FCoE Specification	000:00:00.000_016_678	VLAN 3 ; Cisco Systems:A3:D3:C0	VLAN 3 ; 0e:fc:00:2b:00:23	2,1,1 / 2,1,2	GigE	3007		

Figure 5-32. Expert Shows PFC Pause Request Released from the Target to the Switch Port

Here are a few observations on the virtual system:

- Throughput performance is less stable when running applications with virtual machines than without.
- FCoE traffic throughput decreased because PFC paused FCoE traffic. In contrast, the iSCSI application gained bandwidth while FCoE traffic was paused. These changes in throughput might have been caused by increased virtual OS processing overhead, competition for CPU time, or transmission scheduling conflicts between the virtual machines.
- iSCSI throughput increases sharply when FCoE traffic decreases in the absence of PFC pauses (Figure 5-31).

Finally, compare the results from the virtual environment with the results from the multiple host environment.

In summary, multiple factors may determine the behavior of virtual-server-based applications in the unified network environment. The internal application balancing algorithm (either among virtual servers or within the unified storage) plays the key role, and could interfere with the DCB-based bandwidth management during congestion.

A Hardware and Software

Cisco Unified Fabric Switch

The Cisco Nexus 5000 Series switch enables a high-performance, standards-based, Ethernet unified fabric. The platform consolidates separate LAN, SAN, and server cluster environments into a single physical fabric while preserving existing operational models. The Cisco Nexus 5000 Series switch provides an enhanced Ethernet topology by leveraging data center bridging features, which include priority flow control (PFC) for a lossless fabric and enhanced transmission selection (ETS) for bandwidth management. These Ethernet enhancements allow technologies like Fibre Channel over Ethernet, and allow consolidation of I/O without compromise. A unified fabric enables increased bandwidth usage, less cabling, fewer adapters, and less network equipment. The benefits are: reduced power and cooling requirements, significant cost savings in infrastructure software and hardware, and reduced infrastructure management costs.

The Nexus 5000 Series switch uses cut-through architecture, supports line-rate 10 Gigabit Ethernet on all ports, and maintains consistently low latency (independent of packet size and enabled services). In addition, the Nexus 5000 Series switch supports a set of network technologies known collectively as IEEE data center bridging (DCB) that increases the reliability, efficiency, and scalability of Ethernet networks. These features enable support for multiple traffic classes over a lossless Ethernet fabric, thus enabling consolidation of LAN, SAN, and cluster environments. The ability to connect FCoE to native Fibre Channel protects existing storage system investments, while dramatically simplifying in-rack cabling.

For more information about the Cisco Nexus 5000 Series switch, visit http://www.cisco.com/en/US/products/ps9670/index.html.

For assistance using the JDSU tool, contact technical support at 1-866-594-2557.

JDSU (formerly Finisar) Equipment and Software

JDSU Xgig

JDSU Xgig® is a unified, integrated platform employing a unique chassis and blade architecture to provide users with the utmost in scalability and flexibility. Various blades support a wide range of protocols and can be easily configured to act as a protocol Analyzer, Jammer, bit error rate tester (BERT), traffic generator, and Load Tester, all without changing hardware.

Xgig can be placed either directly in-line on a link, or connected using the JDSU family of copper or optical test access points (TAPs). Additionally, multiple Xgig chassis can be cascaded together to provide up to 64 time-synchronized analysis/test ports across multiple protocols, enabling correlation of traffic across several devices and network domains.

- The Xgig Analyzer provides complete visibility into network behaviors with 100 percent capture at full line rate. Xgig is the only protocol Analyzer supporting multiple protocols, all within the same chassis.

  For more information about the Xgig Analyzer, visit http://www.jdsu.com/products/communications-test-measurement/products/a-z-product-list/xgig-protocol-analyzer-family-overview.html.

- The Xgig Jammer manipulates live network traffic to simulate errors in real time, enabling users to verify the performance of error recovery processes.

- The Xgig Load Tester enables developers to quickly and easily verify data integrity, monitor network performance, and identify a wide range of problems across even the most complex network topologies that would be too difficult to troubleshoot with an analyzer alone.

JDSU Medusa Labs Test Tool Suite

The Medusa Labs Test Tool (MLTT) Suite 3.0 provides a comprehensive set of benchmarking, data integrity, and stress test tools that uncover and eliminate data corruption errors, undesirable device and system data pattern sensitivities, I/O timeouts, I/O losses, and system lockup.

For more information about the Medusa Labs Test Tool Suite, visit http://www.jdsu.com/products/communications-test-measurement/products/a-z-product-list/medusa-labs-test-tools-suite.html.

QLogic QLE8142 Converged Network Adapter

The QLogic QLE8142 Converged Network Adapter is a single chip, fully-offloaded FCoE initiator, operating in both virtual and non-virtual environments, running over an Enhanced Ethernet fabric. The QLE8142 adapter initiator boosts system performance with 10Gbps speed and full hardware offload for FCoE protocol processing. Cutting-edge 10Gbps bandwidth eliminates performance bottlenecks in the I/O path with a 10X data rate improvement over existing 1Gbps Ethernet solutions. In addition, full hardware offload for FCoE protocol processing reduces system CPU usage for I/O operations, which leads to faster application performance and greater consolidation in virtualized systems.

QLogic 5800V Series Fibre Channel Switch

The QLogic SB5800V Series Fibre Channel switch supports 2Gb, 4Gb, and 8Gb devices and optics, while providing 10Gb inter-switch links that are can be upgraded to 20Gb when workload demands increase. The QLogic 5800V Series switch provides intuitive installation and configuration wizards, stack and SNMP administration, standard adaptive trunking, and unparalleled upgrade flexibility. Multi-switch implementations require up to 50 percent fewer switches to achieve the same number of device ports. The result is that system costs can grow as needed rather than by over-provisioning Fibre Channel infrastructure upon the initial hardware and software acquisition.

For more information about QLogic's products, visit http://www.qlogic.com/PRODUCTS/Pages/products_landingpage.aspx.

B Data Center Bridging Technology

The following descriptions of Enhanced Ethernet were taken from *Ethernet: The Converged Network Ethernet Alliance Demonstration* at SC'09, published by the Ethernet Alliance November, 2009.

Data Center Bridging (DCB)

For Ethernet to carry LAN, SAN and IPC traffic together and achieve network convergence, some necessary enhancements are required. These enhancement protocols are summarized as data center bridging (DCB) protocols, also referred to as Enhanced Ethernet (EE), which are defined by the IEEE 802.1 data center bridging task group. A converged Ethernet network is built based on the following DCB protocols:

- DCBX and ETS
- Priority Flow Control
- Fibre Channel Over Ethernet
- iSCSI

DCBX and ETS

Existing Ethernet standards cannot control and manage the allocation of network bandwidth to different network traffic sources and types (traffic differentiation). Neither can existing standards allow prioritizing of bandwidth usage across these sources and traffic types. Data center managers must either over-provision network bandwidth for peak loads, accept customer complaints during these periods, or manage traffic on the source side by limiting the amount of non-priority traffic entering the network.

Overcoming these limitations is a key to enabling Ethernet as the foundation for true converged data center networks supporting LAN, storage, and inter-processor communications.

Enhanced Transmission Selection (ETS) protocol addresses the bandwidth allocation issues among various traffic classes to maximize bandwidth usage. The IEEE 802.1Qaz standard specifies the protocol to support allocation of bandwidth amongst priority groups. ETS allows each node to control bandwidth per priority group. When the actual load in a priority group does not use its allocated bandwidth, ETS allows other priority groups to use the available bandwidth. The bandwidth-allocation priorities allow the sharing of bandwidth between traffic loads, while satisfying the strict priority mechanisms already defined in IEEE 802.1Q that require minimum latency.

Bandwidth allocation is achieved as part of a negotiation process with link peers—this is called DCB Capability eXchange protocol (DCBX). It provides a mechanism for Ethernet devices (bridges, end stations) to detect the DCB capability of a peer device. It also allows configuration and distribution of ETS parameters from one node to another.

ETS and DCBX simplify the management of DCB nodes significantly, especially when deployed end-to-end in a converged data center. The DCBX protocol uses Link Layer Discovery Protocol (LLDP) defined by IEEE 802.1AB to exchange and discover DCB capabilities.

Priority Flow Control

A fundamental requirement for a high performance storage network is guaranteed data delivery. This requirement must be satisfied to transport critical storage data on a converged Ethernet network with minimum latency. Another critical enhancement to conventional Ethernet is lossless Ethernet. IEEE 802.3X PAUSE defines how to pause link traffic at a congestion point to avoid packet drop. IEEE 802.1Qbb defines Priority Flow Control (PFC), which is based on IEEE 802.3X PAUSE and provides greater control of traffic flow. PFC eliminates lost frames caused by congestion. PFC enables the pausing of less sensitive data classes, while not affecting traditional LAN protocols operating through different priority classes.

Figure B-1 shows how PFC works in a converged traffic scenario.

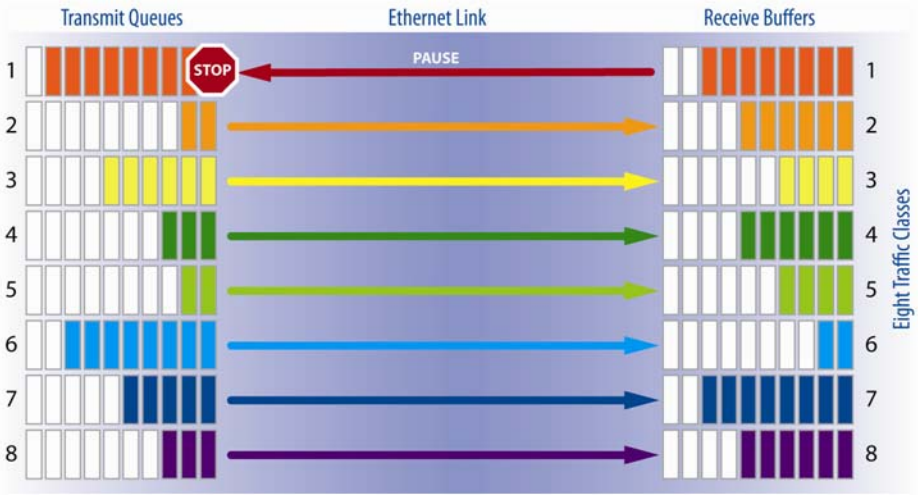


Figure B-1. Priority Flow Control

Fibre Channel Over Ethernet

FCoE is an ANSI T11 standard for the encapsulation of a complete Fibre Channel frame into an Ethernet frame. The resulting Ethernet frame is transported over Enhanced Ethernet networks as shown in Figure B-2. Compared to other mapping technologies, FCoE has the least mapping overhead and maintains the same constructs as native Fibre Channel, thus operating with native Fibre Channel management software. FCoE is based on lossless Ethernet to enable buffer-to-buffer credit management and flow control of Fibre Channel packets.

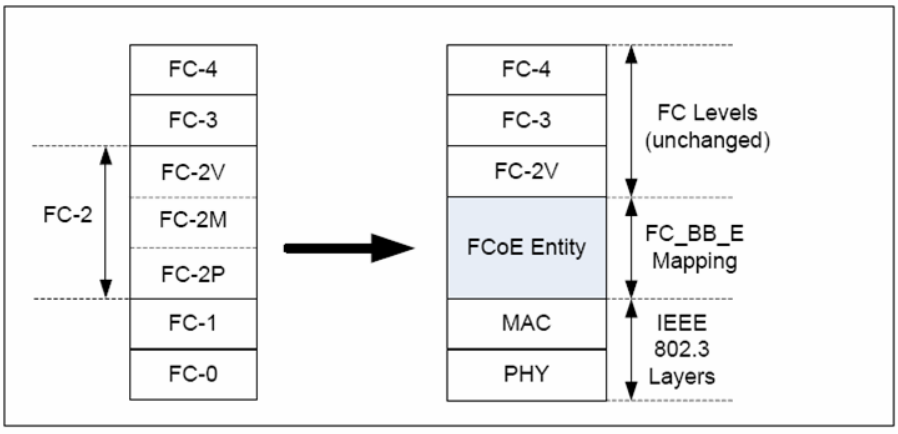


Figure B-2. FCoE Mapping Illustration (Source: FC-BB-5 Rev 2.0)

iSCSI

The Internet Small Computer Systems Interface (iSCSI) is a SCSI mass storage transport that operates between the Transport Control Protocol (TCP) and the SCSI Protocol Layers. The iSCSI protocol is defined in RFC 3720 [iSCSI], which was finalized by the Internet Engineering Task Force (IETF) in April, 2004. A TCP/IP connection ties the iSCSI initiator and target session components together. Network portals identified by their IP address and TCP port numbers define the endpoints of a connection. iSCSI is by nature, a lossless storage network because inherent in iSCSI's design is recovery from dropped packets on over-subscribed, heavy network traffic patterns. iSCSI relies on TCP/IP (or SCTP) for the retransmission of dropped Ethernet frames.

C References

Fibre Channel over Ethernet Design Guide, Cisco and QLogic (2010), QLogic adapters and Cisco Nexus 5000 Series switches, Cisco document number C11-569320-01. Downloaded from http://www.qlogic.com/SiteCollectionDocuments/Education_and_Resource/whitepapers/whitepaper2/QLogic_Cisco_FCoE_Design_Guide.pdf

Ethernet: The Converged Network Ethernet Alliance Demonstration at Sc'09, Ethernet Alliance. (2009), retrieved from http://www.ethernetalliance.org/files/static_page_files/281AD8C4-1D09-3519-AD7AD835AD525E36/SC09%20white%20paper.pdf

Unified Data Center Fabric Primer: Fcoe And Data Center Bridging, Martin, D. (2010). SearchNetworking.com, retrieved from http://searchnetworking.techtarget.com/tip/0,289483,sid7_gci1378613,00.html

Unified fabric: Data Center Bridging And Fcoe Implementation, Martin, D. (2010). SearchNetworking.com, retrieved from http://searchnetworking.techtarget.com/tip/0,289483,sid7_gci1379716_mem1,00.html?ShortReg=1&mboxConv=searchNetworking_RegActivate_Submit&

Index

A

adapters 4-2, 4-3
architecture 2-1, 3-1
audience vii

C

cabling 4-1, 4-5
configuration 4-1
connectivity 4-10
conventions viii
converged network 1-1, 4-2

D

data center bridging exchange protocol vii
design guide 4-5
drivers 4-3

E

enhanced transmission selection vii, 5-1, 5-18
equipment 3-3
  connectivity 4-10
ETS - see enhanced transmission selection
exchange completion time 5-25

F

FCoE
  function validation 5-12
  switch 4-6
FCoE initialization protocol validation 5-6
Fibre Channel forwarder 2-2
Fibre Channel switches 4-5
FIP - see FCoE initialization protocol

H

hardware platforms 4-1

I

I/O test validation 5-12, 5-26
installation 4-1
iSCSI function validation 5-26

L

latency monitor 5-25
lossless definition 1-1
lossless link validation 5-15

M

management practices 2-1

N

NETtrack data center 2-1
Nexus switch 4-6

O

operating systems 4-1

organizational ownership 2-1

P

PFC - see priority flow control
planning 2-1
priority flow control vii, 5-1, 5-15
process summary 3-1

Q

QLogic and Cisco FCoE Design Guide 4-5

R

related materials vii

S

SANsurfer FC HBA Manager 4-4
server bus interface 4-1
storage 4-1, 4-9
switch
  FCoE 4-5, 4-6
  Fibre Channel 4-5
  Nexus 4-6
  types 4-1

T

test architecture 2-1

U

unified data center 1-1
unified fabric
  certify 1-1
  components 2-1
  Fibre Channel connections 2-2

V

virtualization verification 5-29

Corporate Headquarters QLogic Corporation 26650 Aliso Viejo Parkway Aliso Viejo, CA 92656 949.389.6000 **www.qlogic.com**

International Offices UK | Ireland | Germany | India | Japan | China | Hong Kong | Singapore | Taiwan

© 2010 QLogic Corporation. Specifications are subject to change without notice. All rights reserved worldwide. QLogic, the QLogic logo, and SANsurfer are registered trademarks of QLogic Corporation. Cisco and Cisco Nexus are trademarks or registered trademarks of Cisco Systems, Inc. Dell and PowerEdge are registered trademarks of Dell Inc. VMware and ESX are trademarks or registered trademarks of VMware Inc. JDSU, Finisar, and Xgig are trademarks or registered trademarks of JDS Uniphase Corporation. Microsoft, Windows, and Windows Server are registered trademarks of Microsoft Corporation. NetApp is a registered trademark of Network Appliance, Inc. HDS is a registered trademark of Hitachi, Ltd. and/or its affiliates. SPARC is a registered trademark of SPARC International, Inc. in the USA and other countries. PowerPC is a registered trademark of International Business Machines Corporation. Red Hat is a registered trademark of Red Hat, Inc. Novell is a registered trademark of Novell, Inc. Solaris and OpenSolaris are trademarks or registered trademarks of Sun Microsystems, Inc. All other brand and product names are trademarks or registered trademarks of their respective owners. Information supplied by QLogic Corporation is believed to be accurate and reliable. QLogic Corporation assumes no responsibility for any errors in this brochure. QLogic Corporation reserves the right, without notice, to make changes in product design or specifications.